

*This book is dedicated
to the memory of
Kenneth E. Easterling*

Phase Transformations in Metals and Alloys

SECOND EDITION

D.A. Porter
*Rautaruukki Oy
Research Centre
Raahe
Finland*

K.E. Easterling
*Formerly School of Engineering
University of Exeter
Exeter
UK*



CHAPMAN & HALL

London · Weinheim · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK

Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK

Chapman & Hall GmbH, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA, 115 Fifth Avenue, New York, NY 10003, USA

Chapman & Hall Japan, ITP-Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1981

Reprinted 1982, 1983, 1984, 1986, 1987, 1988, 1989, 1990, 1992

Second edition 1992

Reprinted 1993 (twice), 1995, 1996

© 1981, 1992 D.A. Porter and K.E. Easterling

Typeset in 10/12pt Times by Best-set Typesetter Ltd, Hong Kong

Printed in Great Britain by T.J. Press (Padstow) Ltd, Cornwall

ISBN 0 412 45030 5

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright Designs and Patents Act, 1988, this publication may not be reproduced, stored, or transmitted, in any form or by any means, without the prior permission in writing of the publishers, or in the case of reprographic reproduction only in accordance with the terms of the licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to the publishers at the London address printed on this page.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

A Catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data available

Printed on permanent acid-free text paper, manufactured in accordance with ANSI/NISO Z39.48-1992 and ANSI/NISO Z39.48-1984 (Permanence of Paper).

Contents

<i>Preface to the second edition</i>	xi
<i>Preface to the first edition</i>	xii
1 Thermodynamics and Phase Diagrams	1
1.1 Equilibrium	1
1.2 Single Component Systems	4
1.2.1 Gibbs Free Energy as a Function of Temperature	4
1.2.2 Pressure Effects	7
1.2.3 The Driving Force for Solidification	10
1.3 Binary Solutions	11
1.3.1 The Gibbs Free Energy of Binary Solutions	11
1.3.2 Ideal Solutions	13
1.3.3 Chemical Potential	16
1.3.4 Regular Solutions	18
1.3.5 Activity	21
1.3.6 Real Solutions	23
1.3.7 Ordered Phases	24
1.3.8 Intermediate Phases	26
1.4 Equilibrium in Heterogeneous Systems	28
1.5 Binary Phase Diagrams	33
1.5.1 A Simple Phase Diagram	33
1.5.2 Systems with a Miscibility Gap	33
1.5.3 Ordered Alloys	35
1.5.4 Simple Eutectic Systems	36
1.5.5 Phase Diagrams Containing Intermediate Phases	36
1.5.6 The Gibbs Phase Rule	36
1.5.7 The Effect of Temperature on Solid Solubility	41
1.5.8 Equilibrium Vacancy Concentration	43
1.6 The Influence of Interfaces on Equilibrium	44
1.7 Ternary Equilibrium	49
1.8 Additional Thermodynamic Relationships for Binary Solutions	52
1.9 The Kinetics of Phase Transformations	55
References	56
Further Reading	56
Exercises	57

2 Diffusion	60
2.1 Atomic Mechanisms of Diffusion	61
2.2 Interstitial Diffusion	63
2.2.1 Interstitial Diffusion as a Random Jump Process	63
2.2.2 Effect of Temperature—Thermal Activation	66
2.2.3 Steady-State Diffusion	69
2.2.4 Nonsteady-State Diffusion	69
2.2.5 Solutions to the Diffusion Equation	71
Homogenization	71
The Carburization of Steel	73
2.3 Substitutional Diffusion	75
2.3.1 Self-Diffusion	75
2.3.2 Vacancy Diffusion	79
2.3.3 Diffusion in Substitutional Alloys	82
2.3.4 Diffusion in Dilute Substitutional Alloys	91
2.4 Atomic Mobility	92
2.5 Tracer Diffusion in Binary Alloys	94
2.6 Diffusion in Ternary Alloys	96
2.7 High-Diffusivity Paths	98
2.7.1 Diffusion along Grain Boundaries and Free Surfaces	98
2.7.2 Diffusion along Dislocations	102
2.8 Diffusion in Multiphase Binary Systems	103
References	106
Further Reading	106
Exercises	106
3 Crystal Interfaces and Microstructure	110
3.1 Interfacial Free Energy	110
3.2 Solid/Vapour Interfaces	112
3.3 Boundaries in Single-Phase Solids	116
3.3.1 Low-Angle and High-Angle Boundaries	116
3.3.2 Special High-Angle Grain Boundaries	122
3.3.3 Equilibrium in Polycrystalline Materials	124
3.3.4 Thermally Activated Migration of Grain Boundaries	130
3.3.5 The Kinetics of Grain Growth	139
3.4 Interphase Interfaces in Solids	142
3.4.1 Interface Coherence	143
Fully Coherent Interfaces	143
Semicoherent Interfaces	145
Incoherent Interfaces	147
Complex Semicoherent Interfaces	148
3.4.2 Second-Phase Shape: Interfacial Energy Effects	149
Fully Coherent Precipitates	149
Partially Coherent Precipitates	151
Incoherent Precipitates	152
Precipitates on Grain Boundaries	153

3.4.3 Second-Phase Shape: Misfit Strain Effects	154
Fully Coherent Precipitates	154
Incoherent Inclusions	158
Plate-Like Precipitates	160
3.4.4 Coherency Loss	160
3.4.5 Glissile Interfaces	163
3.4.6 Solid/Liquid Interfaces	168
3.5 Interface Migration	171
3.5.1 Diffusion-Controlled and Interface-Controlled Growth	175
References	180
Further Reading	182
Exercises	182
4 Solidification	185
4.1 Nucleation in Pure Metals	185
4.1.1 Homogeneous Nucleation	186
4.1.2 The Homogeneous Nucleation Rate	190
4.1.3 Heterogeneous Nucleation	192
4.1.4 Nucleation of Melting	197
4.2 Growth of a Pure Solid	197
4.2.1 Continuous Growth	198
4.2.2 Lateral Growth	198
Surface Nucleation	200
Spiral Growth	201
Growth from Twin Intersections	202
4.2.3 Heat Flow and Interface Stability	203
4.3 Alloy Solidification	207
4.3.1 Solidification of Single-Phase Alloys	208
Equilibrium Solidification	208
No Diffusion in Solid, Perfect Mixing in Liquid	208
No Diffusion in Solid, Diffusional Mixing in Liquid	212
Cellular and Dendritic Solidification	214
4.3.2 Eutectic Solidification	222
Growth of Lamellar Eutectics	223
4.3.3 Off-Eutectic Alloys	229
4.3.4 Peritectic Solidification	231
4.4 Solidification of Ingots and Castings	233
4.4.1 Ingot Structure	233
Chill Zone	234
Columnar Zone	235
Equiaxed Zone	236
Shrinkage Effects	236
4.4.2 Segregation in Ingots and Castings	237
4.4.3 Continuous Casting	238
Heat Flow in Welding and Continuous Casting	239

4.5	Solidification of Fusion Welds	243
	Influence of Welding Speed	245
	Geometry of Crystal Growth	248
4.6	Solidification during Quenching from the Melt	249
4.7	Case Studies of some Practical Castings and Welds	249
	4.7.1 Casting of Carbon and Low-Alloy Steels	249
	4.7.2 Casting of High-Speed Steels	251
	4.7.3 Stainless Steel Weld Metal	256
	References	259
	Further Reading	260
	Exercises	260
5	Diffusional Transformations in Solids	263
5.1	Homogeneous Nucleation in Solids	265
5.2	Heterogeneous Nucleation	271
	Nucleation on Grain Boundaries	271
	Dislocations	274
	Excess Vacancies	275
	5.2.1 Rate of Heterogeneous Nucleation	276
5.3	Precipitate Growth	279
	5.3.1 Growth behind Planar Incoherent Interfaces	279
	5.3.2 Diffusion-Controlled Lengthening of Plates or Needles	283
	5.3.3 Thickening of Plate-like Precipitates	285
5.4	Overall Transformation Kinetics—TTT Diagrams	287
5.5	Precipitation in Age-Hardening Alloys	291
	5.5.1 Precipitation in Aluminium—Copper Alloys	291
	GP Zones	291
	Transition Phases	292
	5.5.2 Precipitation in Aluminium—Silver Alloys	302
	5.5.3 Quenched-in Vacancies	303
	5.5.4 Age Hardening	304
	5.5.5 Spinodal Decomposition	308
	5.5.6 Particle Coarsening	314
	Low γ	316
	Low X_c	316
	Low D	317
5.6	The Precipitation of Ferrite from Austenite	317
5.7	Cellular Precipitation	322
5.8	Eutectoid Transformations	326
	5.8.1 The Pearlite Reaction in Fe—C Alloys	326
	Nucleation of Pearlite	327
	Pearlite Growth	330
	Pearlite in Off-Eutectoid Fe—C Alloys	333
	5.8.2 The Bainite Transformation	334

	Upper Bainite	334
	Lower Bainite	337
	Transformation Shears	337
5.8.3	The Effect of Alloying Elements on Hardenability	338
5.8.4	Continuous Cooling Diagrams	344
5.8.5	Fibrous and Interphase Precipitation in Alloy Steels	349
5.9	Massive Transformations	349
5.10	Ordering Transformations	358
5.11	Case Studies	366
	5.11.1 Titanium Forging Alloys	366
	5.11.2 The Weldability of Low-Carbon and Microalloyed	
	Rolled Steels	372
	References	377
	Further Reading	378
	Exercises	379
6	Diffusionless Transformations	382
6.1	Characteristics of Diffusionless Transformations	383
	6.1.1 The Solid Solution of Carbon in Iron	385
6.2	Martensite Crystallography	389
	6.2.1 The Bain Model of the fcc $\rightarrow$ bcc Transformation	391
	6.2.2 Comparison of Crystallographic Theory with	
	Experimental Results	396
6.3	Theories of Martensite Nucleation	397
	6.3.1 Formation of Coherent Nuclei of Martensite	398
	6.3.2 Role of Dislocations in Martensite Nucleation	401
	6.3.3 Dislocation Strain Energy Assisted Transformation	406
6.4	Martensite Growth	409
	6.4.1 Growth of Lath Martensite	410
	6.4.2 Plate Martensite	412
	6.4.3 Stabilization	415
	6.4.4 Effect of External Stresses	415
	6.4.5 Role of Grain Size	416
6.5	Pre-martensite Phenomena	416
6.6	Tempering of Ferrous Martensites	417
	Carbon Segregation	420
	ϵ -Carbide	421
	Cementite	422
	Alloy Carbides	422
	Effect of Retained Austenite	426
	Recovery, Recrystallization and Grain Growth	426
	Temper Embrittlement	427
6.7	Case Studies	428
	6.7.1 Carbon and Low-Alloy Quenched and Tempered Steels	428
	6.7.2 Controlled Transformation Steels	430

6.7.3 The 'Shape-Memory' Metal: Nitinol	431
References	437
Further Reading	439
Exercises	439
Solutions to exercises	441
Compiled by John C. Ion	
Index	510

Preface to the second edition

In the ten years since this book was first published there have been many new developments in the metallurgical field. Rapidly solidified metals and glasses have come of age; new Al-Li alloys are now used in modern aircraft; microalloyed (structural) and high purity (pipeline) steels have become more sophisticated; radically new oxide-dispersed steels have appeared; a number of new memory metals have been developed; the list could go on. In spite of this, the underlying principles governing all of these developments have obviously not changed over the years. This is really the strength of the present text book. From the beginning we aimed to spell out these principles in a nice, readable way, and one in which undergraduates could appreciate and be capable of developing for themselves. The present text is thus deliberately little changed from the original. We have, however, hopefully corrected any errors, expanded the lists of further reading, and perhaps, most importantly, included a complete set of solutions to exercises. We hope that the revised edition continues to be enjoyed and appreciated in the many Schools of Metallurgy, Materials Science and Engineering Materials we know to be using our text throughout the world.

In completing this revised edition we are grateful to the many people, students and professors alike, who have written to us over the last decade. Particular thanks are due to Dr Wen-Bin Li (University of Luleå) for using a fine tooth-comb in bringing out both obvious and less obvious errors in the original text. There remain, (inevitably), a few 'points of contention' concerning our description of certain phenomena, as raised by some of our correspondents, but there is nothing unhealthy about that. We should finally like to thank Dr John Ion (University of Lappeenranta, Finland) for his help in compiling the Solutions to Exercises chapter.

David Porter and Kenneth Easterling
September 1991

Preface to the first edition

This book is written as an undergraduate course in phase transformations for final year students specializing in metallurgy, materials science or engineering materials. It should also be useful for research students interested in revising their knowledge of the subject. The book is based on lectures originally given by the authors at the University of Luleå for engineering students specializing in engineering materials. Surprisingly we found no modern treatments of this important subject in a form suitable for a course book, the most recent probably being P.G. Shewmon's *Transformations in Metals* (McGraw-Hill, 1969). There have, however, been some notable developments in the subject over the last decade, particularly in studies of interfaces between phases and interface migration, as well as the kinetics of precipitate growth and the stability of precipitates. There have also been a number of important new practical developments based on phase transformations, including the introduction of TRIP steels (transformation induced by plastic deformation), directionally aligned eutectic composites, and sophisticated new structural steels with superior weldability and forming properties, to mention just a few. In addition, continuous casting and high speed, high energy fusion welding have emerged strongly in recent years as important production applications of solidification. It was the objective of this course to present a treatment of phase transformations in which these and other new developments could be explained in terms of the basic principles of thermodynamics and atomic mechanisms.

The book is effectively in two parts. Chapters 1-3 contain the background material necessary for understanding phase transformations: thermodynamics, kinetic, diffusion theory and the structure and properties of interfaces. Chapters 4-6 deal with specific transformations: solidification, diffusional transformations in solids and diffusionless transformations. At the end of the chapters on solidification, diffusion-controlled transformations and martensite, we give a few selected case studies of engineering alloys to illustrate some of the principles discussed earlier. In this way, we hope that the text will provide a useful link between theory and the practical reality. It should be stated that we found it necessary to give this course in conjunction with a number of practical laboratory exercises and worked examples. Sets of problems are also included at the end of each chapter of the book.

In developing this course and writing the text we have had continuous

support and encouragement of our colleagues and students in the Department of Engineering Materials. Particular thanks are due to Agneta Engfors for her patience and skill in typing the manuscript as well as assisting with the editing.

David Porter and Kenneth Easterling
February 1980

Thermodynamics and Phase Diagrams

This chapter deals with some of the basic thermodynamic concepts that are required for a more fundamental appreciation of phase diagrams and phase transformations. It is assumed that the student is already acquainted with elementary thermodynamics and only a summary of the most important results as regards phase transformations will be given here. Fuller treatment can be found in the books listed in the bibliography at the end of this chapter.

The main use of thermodynamics in physical metallurgy is to allow the prediction of whether an alloy is in equilibrium. In considering phase transformations we are always concerned with changes towards equilibrium, and thermodynamics is therefore a very powerful tool. It should be noted, however, that the rate at which equilibrium is reached cannot be determined by thermodynamics alone, as will become apparent in later chapters.

1.1 Equilibrium

It is useful to begin this chapter on thermodynamics by defining a few of the terms that will be frequently used. In the study of phase transformations we will be dealing with the changes that can occur within a given *system*, e.g. an alloy that can exist as a mixture of one or more phases. A *phase* can be defined as a portion of the system whose properties and composition are homogeneous and which is physically distinct from other parts of the system. The *components* of a given system are the different elements or chemical compounds which make up the system, and the composition of a phase or the system can be described by giving the relative amounts of each component.

The study of phase transformations, as the name suggests, is concerned with how one or more phases in an alloy (the system) change into a new phase or mixture of phases. The reason why a transformation occurs at all is because the initial state of the alloy is unstable relative to the final state. But how is phase stability measured? The answer to this question is provided by thermodynamics. For transformations that occur at constant temperature and pressure the relative stability of a system is determined by its *Gibbs free energy* (G).

The Gibbs free energy of a system is defined by the equation

$$G = H - TS \quad (1.1)$$

where H is the enthalpy, T the absolute temperature, and S the entropy of the system. Enthalpy is a measure of the heat content of the system and is given

by

$$H = E + PV \quad (1.2)$$

where E is the internal energy of the system, P the pressure, and V the volume. The internal energy arises from the total kinetic and potential energies of the atoms within the system. Kinetic energy can arise from atomic vibration in solids or liquids and from translational and rotational energies for the atoms and molecules within a liquid or gas; whereas potential energy arises from the interactions, or bonds, between the atoms within the system. If a transformation or reaction occurs the heat that is absorbed or evolved will depend on the change in the internal energy of the system. However it will also depend on changes in the volume of the system and the term PV takes this into account, so that at *constant pressure* the heat absorbed or evolved is given by the change in H . When dealing with condensed phases (solids and liquids) the PV term is usually very small in comparison to E , that is $H \approx E$. This approximation will be made frequently in the treatments given in this book. The other function that appears in the expression for G is entropy (S) which is a measure of the randomness of the system.

A system is said to be in equilibrium when it is in the most stable state, i.e. shows no desire to change *ad infinitum*. An important consequence of the laws of classical thermodynamics is that at constant temperature and pressure a closed system (i.e. one of fixed mass and composition) will be in stable equilibrium if it has the lowest possible value of the Gibbs free energy, or in mathematical terms

$$dG = 0 \quad (1.3)$$

It can be seen from the definition of G , Equation 1.1, that the state with the highest stability will be that with the best compromise between low enthalpy and high entropy. Thus at low temperatures solid phases are most stable since they have the strongest atomic binding and therefore the lowest internal energy (enthalpy). At high temperatures however the $-TS$ term dominates and phases with more freedom of atom movement, liquids and gases, become most stable. If pressure changes are considered it can be seen from Equation 1.2 that phases with small volumes are favoured by high pressures.

The definition of equilibrium given by Equation 1.3 can be illustrated graphically as follows. If it were possible to evaluate the free energy of a given system for all conceivable configurations the stable equilibrium configuration would be found to have the lowest free energy. This is illustrated in Fig. 1.1 where it is imagined that the various atomic configurations can be represented by points along the abscissa. Configuration A would be the stable equilibrium state. At this point small changes in the arrangement of atoms to a first approximation produce no change in G , i.e. Equation 1.3 applies. However there will always be other configurations, e.g. B, which lie at a local minimum in free energy and therefore also satisfy Equation 1.3, but which do not have the lowest possible value of G . Such configurations are called *metastable*

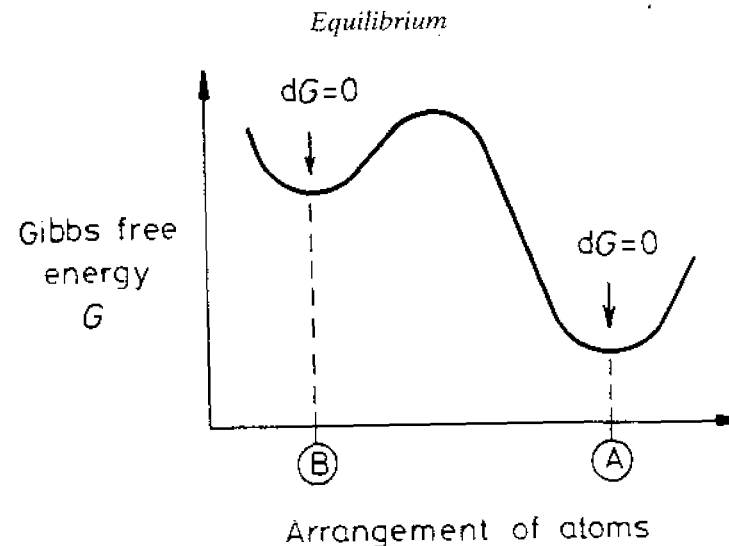


Fig. 1.1 A schematic variation of Gibbs free energy with the arrangement of atoms. Configuration 'A' has the lowest free energy and is therefore the arrangement when the system is at stable equilibrium. Configuration 'B' is a metastable equilibrium.

equilibrium states to distinguish them from the *stable equilibrium* state. The intermediate states for which $dG \neq 0$ are *unstable* and are only ever realized momentarily in practice. If, as the result of thermal fluctuations, the atoms become arranged in an intermediate state they will rapidly rearrange into one of the free energy minima. If by a change of temperature or pressure, for example, a system is moved from a stable to a metastable state it will, given time, transform to the new stable equilibrium state.

Graphite and diamond at room temperature and pressure are examples of stable and metastable equilibrium states. Given time, therefore, all diamond under these conditions will transform to graphite.

Any transformation that results in a decrease in Gibbs free energy is possible. Therefore a necessary criterion for any phase transformation is

$$\Delta G = G_2 - G_1 < 0 \quad (1.4)$$

where G_1 and G_2 are the free energies of the initial and final states respectively. The transformation need not go directly to the stable equilibrium state but can pass through a whole series of intermediate metastable states.

The answer to the question "How fast does a phase transformation occur?" is not provided by classical thermodynamics. Sometimes metastable states can be very short-lived; at other times they can exist almost indefinitely as in the case of diamond at room temperature and pressure. The reason for these differences is the presence of the free energy hump between the metastable and stable states in Fig. 1.1. The study of transformation rates in physical chemistry belongs to the realm of *kinetics*. In general, higher humps or energy barriers lead to slower transformation rates. Kinetics obviously plays a central

role in the study of phase transformations and many examples of kinetic processes will be found throughout this book.

The different thermodynamic functions that have been mentioned in this section can be divided into two types called intensive and extensive properties. The *intensive* properties are those which are independent of the size of the system such as T and P , whereas the *extensive* properties are directly proportional to the quantity of material in the system, e.g. $\bar{V}$, E , H , S and G . The usual way of measuring the size of the system is by the number of moles of material it contains. The extensive properties are then molar quantities, i.e. expressed in units *per mole*. The number of moles of a given component in the system is given by the mass of the component in grams divided by its atomic or molecular weight.

The number of atoms or molecules within 1 mol of material is given by Avogadro's number (N_A) and is 6.023×10^{23} .

1.2 Single Component Systems

Let us begin by dealing with the phase changes that can be induced in a single component by changes in temperature at a fixed pressure, say 1 atm. A single component system could be one containing a pure element or one type of molecule that does not dissociate over the range of temperature of interest. In order to predict the phases that are stable or mixtures that are in equilibrium at different temperatures it is necessary to be able to calculate the variation of G with T .

1.2.1 Gibbs Free Energy as a Function of Temperature

The specific heat of most substances is easily measured and readily available. In general it varies with temperature as shown in Fig. 1.2a. The specific heat is the quantity of heat (in joules) required to raise the temperature of the substance by one degree Kelvin. At constant pressure this is denoted by C_p and is given by

$$C_p = \left(\frac{\partial H}{\partial T} \right)_p \quad (1.5)$$

Therefore the variation of H with T can be obtained from a knowledge of the variation of C_p with T . In considering phase transformations or chemical reactions it is only changes in thermodynamic functions that are of interest. Consequently H can be measured relative to any reference level which is usually done by defining $H = 0$ for a pure element in its most stable state at 298 K (25 °C). The variation of H with T can then be calculated by integrating

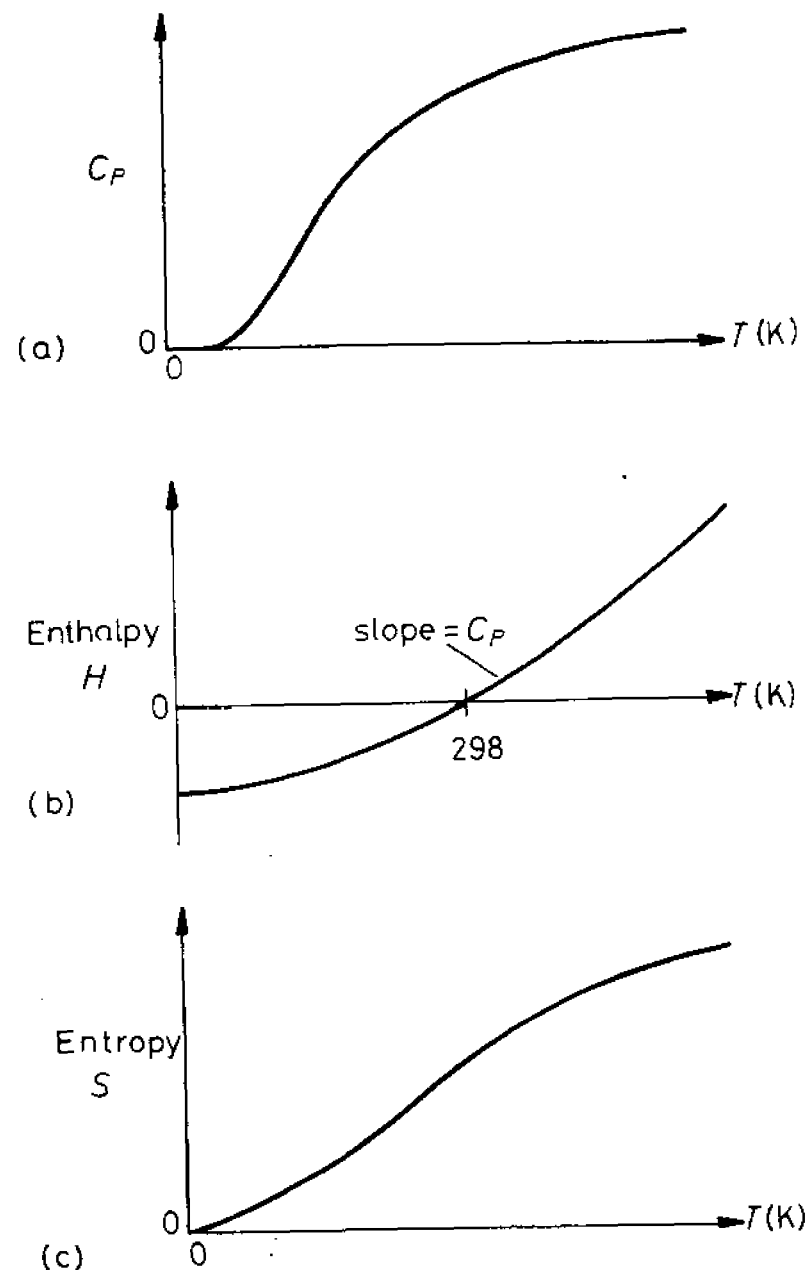


Fig. 1.2 (a) Variation of C_p with temperature, C_p tends to a limit of $\sim 3R$. (b) Variation of enthalpy (H) with absolute temperature for a pure metal. (c) Variation of entropy (S) with absolute temperature.

Equation 1.5, i.e.

$$H = \int_{298}^T C_p dT \quad (1.6)$$

The variation is shown schematically in Fig. 1.2b. The slope of the H - T curve is C_p .

The variation of entropy with temperature can also be derived from the specific heat C_p . From classical thermodynamics

$$\frac{C_p}{T} = \left(\frac{\partial S}{\partial T} \right)_P \quad (1.7)$$

Taking entropy at zero degrees Kelvin as zero, Equation 1.7 can be integrated to give

$$S = \int_0^T \frac{C_p}{T} dT \quad (1.8)$$

as shown in Fig. 1.2c.

Finally the variation of G with temperature shown in Fig. 1.3 is obtained by combining Fig. 1.2b and c using Equation 1.1. When temperature and pressure vary the change in Gibbs free energy can be obtained from the following result of classical thermodynamics: for a system of fixed mass and composition

$$dG = -SdT + VdP \quad (1.9)$$

At constant pressure $dP = 0$ and

$$\left(\frac{\partial G}{\partial T} \right)_P = -S \quad (1.10)$$

This means that G decreases with increasing T at a rate given by $-S$. The relative positions of the free energy curves of solid and liquid phases are illustrated in Fig. 1.4. At all temperatures the liquid has a higher enthalpy (internal energy) than the solid. Therefore at low temperatures $G^L > G^S$. However, the liquid phase has a higher entropy than the solid phase and the Gibbs free energy of the liquid therefore decreases more rapidly with increasing temperature than that of the solid. For temperatures up to T_m the solid phase has the lowest free energy and is therefore the stable equilibrium phase, whereas above T_m the liquid phase is the equilibrium state of the system. At T_m both phases have the same value of G and both solid and liquid can exist in equilibrium. T_m is therefore the equilibrium melting temperature at the pressure concerned.

If a pure component is heated from absolute zero the heat supplied will raise the enthalpy at a rate determined by C_p (solid) along the line ab in Fig. 1.4. Meanwhile the free energy will decrease along ae . At T_m the heat

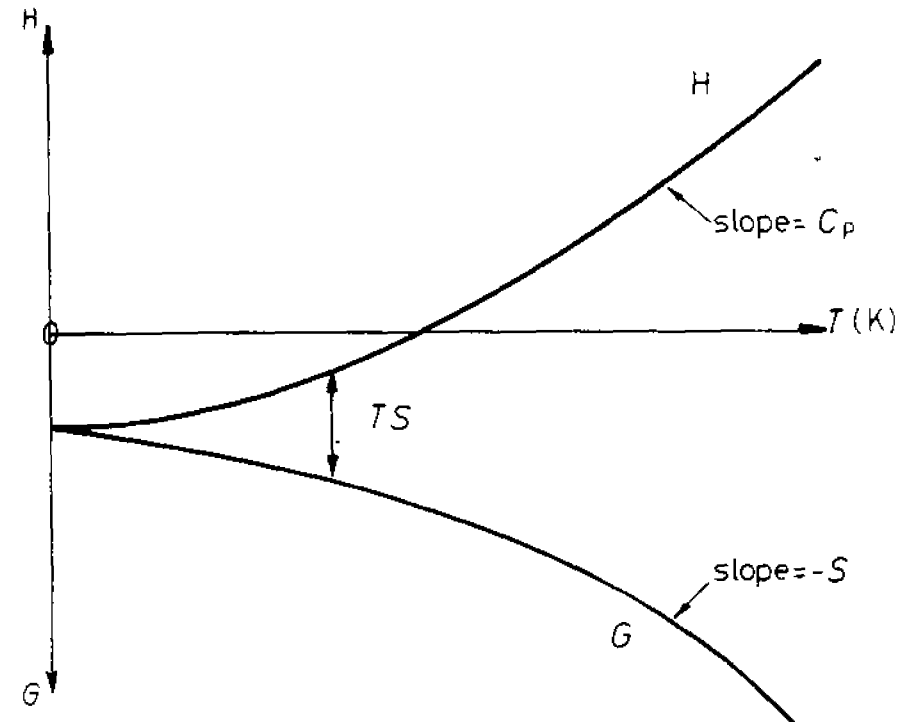


Fig. 1.3 Variation of Gibbs free energy with temperature.

supplied to the system will not raise its temperature but will be used in supplying the latent heat of melting (L) that is required to convert solid into liquid (bc in Fig. 1.4). Note that at T_m the specific heat appears to be infinite since the addition of heat does not appear as an increase in temperature. When all solid has transformed into liquid the enthalpy of the system will follow the line cd while the Gibbs free energy decreases along ef . At still higher temperatures than shown in Fig. 1.4 the free energy of the gas phase (at atmospheric pressure) becomes lower than that of the liquid and the liquid transforms to a gas. If the solid phase can exist in different crystal structures (allotropes or polymorphs) free energy curves can be constructed for each of these phases and the temperature at which they intersect will give the equilibrium temperature for the polymorphic transformation. For example at atmospheric pressure iron can exist as either bcc ferrite below 910°C or fcc austenite above 910°C , and at 910°C both phases can exist in equilibrium.

1.2.2 Pressure Effects

The equilibrium temperatures discussed so far only apply at a specific pressure (1 atm, say). At other pressures the equilibrium temperatures will differ.

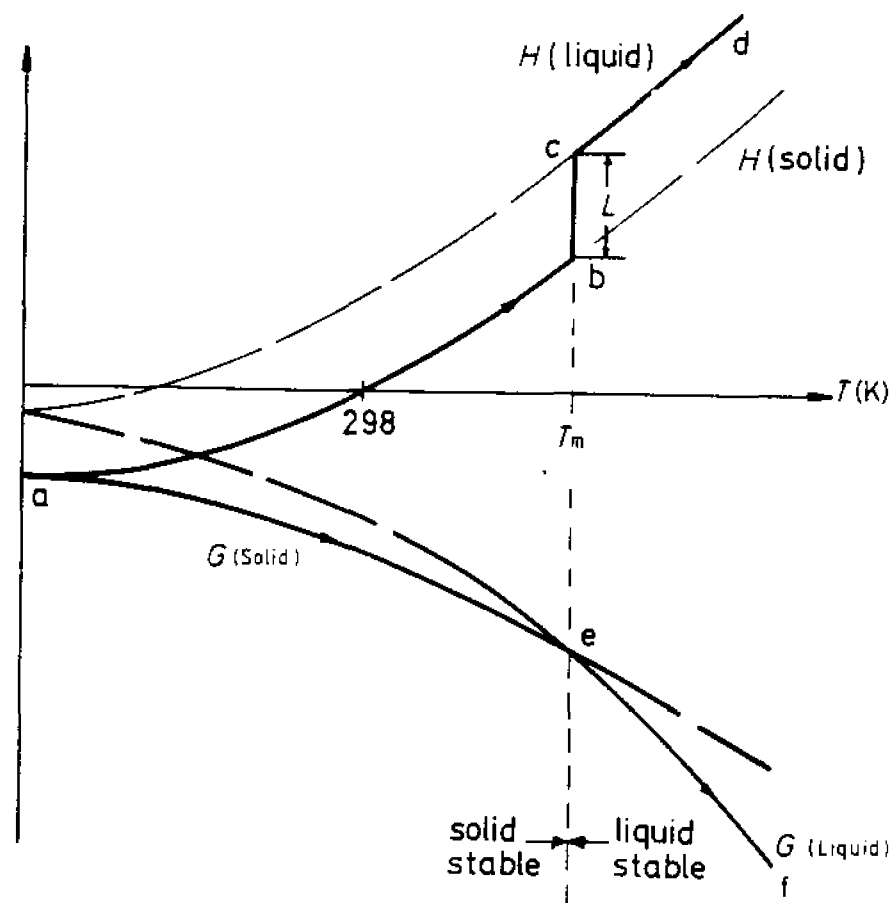


Fig. 1.4 Variation of enthalpy (H) and free energy (G) with temperature for the solid and liquid phases of a pure metal. L is the latent heat of melting, T_m the equilibrium melting temperature.

For example Fig. 1.5 shows the effect of pressure on the equilibrium temperatures for pure iron. Increasing pressure has the effect of depressing the α/γ equilibrium temperature and raising the equilibrium melting temperature. At very high pressures hcp ϵ -Fe becomes stable. The reason for these changes derives from Equation 1.9. At constant temperature the free energy of a phase increases with pressure such that

$$\left(\frac{\partial G}{\partial P}\right)_T = V \quad (1.11)$$

If the two phases in equilibrium have different molar volumes their respective free energies will not increase by the same amount at a given temperature and equilibrium will, therefore, be disturbed by changes in

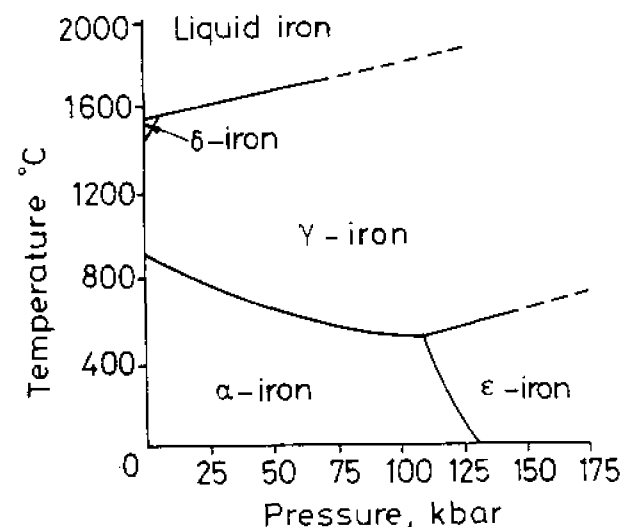


Fig. 1.5 Effect of pressure on the equilibrium phase diagram for pure iron.

pressure. The only way to maintain equilibrium at different pressures is by varying the temperature.

If the two phases in equilibrium are α and β , application of Equation 1.9 to 1 mol of both gives

$$dG^\alpha = V_m^\alpha dP - S^\alpha dT \quad (1.12)$$

$$dG^\beta = V_m^\beta dP - S^\beta dT$$

If α and β are in equilibrium $G^\alpha = G^\beta$ therefore $dG^\alpha = dG^\beta$ and

$$\left(\frac{dP}{dT}\right)_{eq} = \frac{S^\beta - S^\alpha}{V_m^\beta - V_m^\alpha} = \frac{\Delta S}{\Delta V} \quad (1.13)$$

This equation gives the change in temperature dT required to maintain equilibrium between α and β if pressure is increased by dP . The equation can be simplified as follows. From Equation 1.1

$$G^\alpha = H^\alpha - TS^\alpha$$

$$G^\beta = H^\beta - TS^\beta$$

Therefore, putting $\Delta G = G^\beta - G^\alpha$ etc. gives

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

But since at equilibrium $G^\beta = G^\alpha$, $\Delta G = 0$, and

$$\Delta H - T\Delta S = 0$$

Consequently Equation 1.13 becomes

$$\left(\frac{dP}{dT_{eq}}\right) = \frac{\Delta H}{T_{eq}\Delta V} \quad (1.14)$$

which is known as the Clausius–Clapeyron equation. Since close-packed γ -Fe has a smaller molar volume than α -Fe, $\Delta V = V_m^\gamma - V_m^\alpha < 0$ whereas $\Delta H = H^\gamma - H^\alpha > 0$ (for the same reason that a liquid has a higher enthalpy than a solid), so that (dP/dT) is negative, i.e. an increase in pressure lowers the equilibrium transition temperature. On the other hand the δ/L equilibrium temperature is raised with increasing pressure due to the larger molar volume of the liquid phase. It can be seen that the effect of increasing pressure is to increase the area of the phase diagram over which the phase with the smallest molar volume is stable (γ -Fe in Fig. 1.5). It should also be noted that ϵ -Fe has the highest density of the three allotropes, consistent with the slopes of the phase boundaries in the Fe phase diagram.

1.2.3 The Driving Force for Solidification

In dealing with phase transformations we are often concerned with the difference in free energy between two phases at temperatures away from the equilibrium temperature. For example, if a liquid metal is undercooled by ΔT below T_m before it solidifies, solidification will be accompanied by a decrease in free energy ΔG (J mol^{-1}) as shown in Fig. 1.6. This free energy decrease provides the driving force for solidification. The magnitude of this change can be obtained as follows.

The free energies of the liquid and solid at a temperature T are given by

$$G^L = H^L - TS^L$$

$$G^S = H^S - TS^S$$

Therefore at a temperature T

$$\Delta G = \Delta H - T\Delta S \quad (1.15)$$

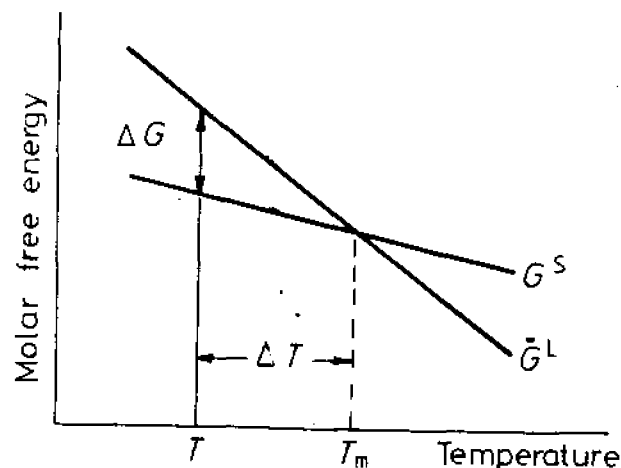


Fig. 1.6 Difference in free energy between liquid and solid close to the melting point. The curvature of the G^S and G^L lines has been ignored.

where

$$\Delta H = H^L - H^S \quad \text{and} \quad \Delta S = S^L - S^S$$

At the equilibrium melting temperature T_m the free energies of solid and liquid are equal, i.e. $\Delta G = 0$. Consequently

$$\Delta G = \Delta H - T_m \Delta S = 0$$

and therefore at T_m

$$\Delta S = \frac{\Delta H}{T_m} = \frac{L}{T_m} \quad (1.16)$$

This is known as the entropy of fusion. It is observed experimentally that the entropy of fusion is a constant $\approx R(8.3 \text{ J mol}^{-1} \text{ K}^{-1})$ for most metals (Richard's rule). This is not unreasonable as metals with high bond strengths can be expected to have high values for both L and T_m .

For small undercoolings (ΔT) the difference in the specific heats of the liquid and solid ($C_p^L - C_p^S$) can be ignored. ΔH and ΔS are therefore approximately independent of temperature. Combining Equations 1.15 and 1.16 thus gives

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

i.e. for small ΔT

$$\Delta G \approx \frac{L\Delta T}{T_m} \quad (1.17)$$

This is a very useful result which will frequently recur in subsequent chapters.

1.3 Binary Solutions

In single component systems all phases have the same composition, and equilibrium simply involves pressure and temperature as variables. In alloys, however, composition is also variable and to understand phase changes in alloys requires an appreciation of how the Gibbs free energy of a given phase depends on composition as well as temperature and pressure. Since the phase transformations described in this book mainly occur at a fixed pressure of 1 atm most attention will be given to changes in composition and temperature. In order to introduce some of the basic concepts of the thermodynamics of alloys a simple physical model for binary solid solutions will be described.

1.3.1 The Gibbs Free Energy of Binary Solutions

The Gibbs free energy of a binary solution of A and B atoms can be calculated from the free energies of pure A and pure B in the following way.

It is assumed that A and B have the same crystal structures in their pure states and can be mixed in any proportions to make a solid solution with the same crystal structure. Imagine that 1 mol of homogeneous solid solution is made by mixing together X_A mol of A and X_B mol of B. Since there is a total of 1 mol of solution

$$X_A + X_B = 1 \quad (1.18)$$

and X_A and X_B are the *mole fractions* of A and B respectively in the alloy. In order to calculate the free energy of the alloy, the mixing can be made in two steps (see Fig. 1.7). These are:

1. bring together X_A mol of pure A and X_B mol of pure B;
2. allow the A and B atoms to mix together to make a homogeneous solid solution.

After step 1 the free energy of the system is given by

$$G_1 = X_A G_A + X_B G_B \quad \text{J mol}^{-1} \quad (1.19)$$

where G_A and G_B are the molar free energies of pure A and pure B at the temperature and pressure of the above experiment. G_1 can be most conveniently represented on a *molar free energy diagram* (Fig. 1.8) in which molar free energy is plotted as a function of X_B or X_A . For all alloy compositions G_1 lies on the straight line between G_A and G_B .

The free energy of the system will not remain constant during the mixing of the A and B atoms and after step 2 the free energy of the solid solution G_2 can

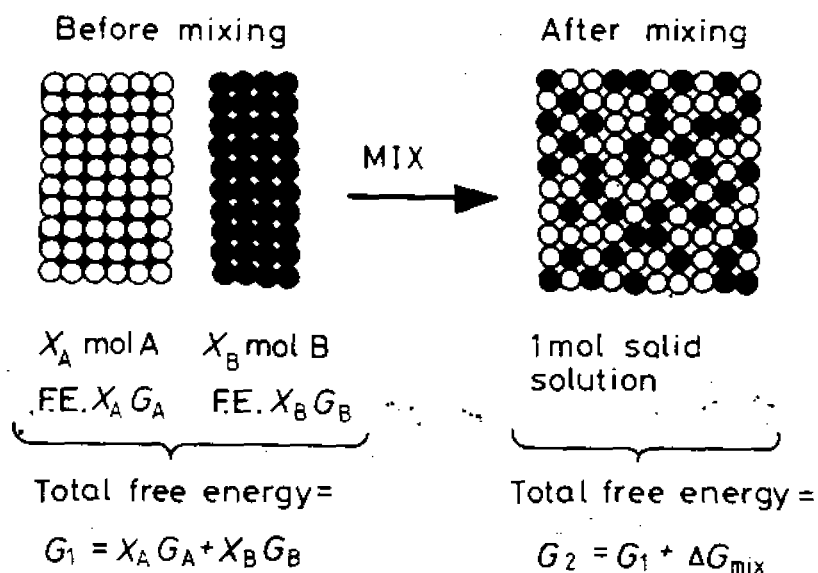


Fig. 1.7 Free energy of mixing.

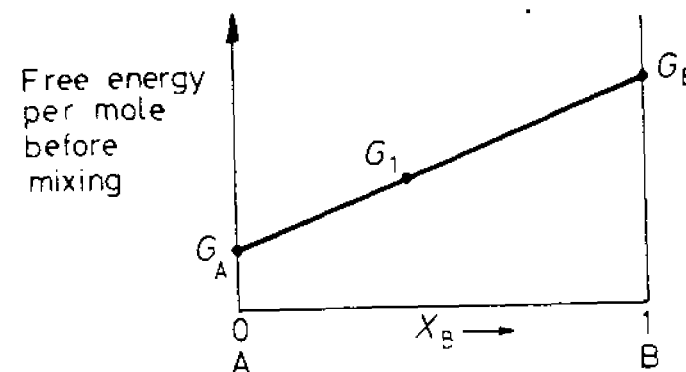


Fig. 1.8 Variation of G_1 (the free energy before mixing) with alloy composition (X_A or X_B).

be expressed as

$$G_2 = G_1 - \Delta G_{\text{mix}} \quad (1.20)$$

where ΔG_{mix} is the change in Gibbs free energy caused by the mixing.

Since

$$G_1 = H_1 - TS_1$$

and

$$G_2 = H_2 - TS_2$$

putting

$$\Delta H_{\text{mix}} = H_2 - H_1$$

and

$$\Delta S_{\text{mix}} = S_2 - S_1$$

gives

$$\Delta G_{\text{mix}} = \Delta H_{\text{mix}} - T\Delta S_{\text{mix}} \quad (1.21)$$

ΔH_{mix} is the heat absorbed or evolved during step 2, i.e. it is the heat of solution, and ignoring volume changes during the process, it represents only the difference in internal energy (E) before and after mixing. ΔS_{mix} is the difference in entropy between the mixed and unmixed states.

1.3.2 Ideal Solutions

The simplest type of mixing to treat first is when $\Delta H_{\text{mix}} = 0$, in which case the resultant solution is said to be *ideal* and the free energy change on mixing is

only due to the change in entropy:

$$\Delta G_{\text{mix}} = -T\Delta S_{\text{mix}} \quad (1.22)$$

In statistical thermodynamics, entropy is quantitatively related to randomness by the Boltzmann equation, i.e.

$$S = k \ln \omega \quad (1.23)$$

where k is Boltzmann's constant and ω is a measure of randomness. There are two contributions to the entropy of a solid solution—a thermal contribution S_{th} and a configurational contribution S_{config} .

In the case of thermal entropy, ω is the number of ways in which the thermal energy of the solid can be divided among the atoms, that is, the total number of ways in which vibrations can be set up in the solid. In solutions, additional randomness exists due to the different ways in which the atoms can be arranged. This gives extra entropy S_{config} for which ω is the number of distinguishable ways of arranging the atoms in the solution.

If there is no volume change or heat change during mixing then the only contribution to ΔS_{mix} is the change in configurational entropy. Before mixing, the A and B atoms are held separately in the system and there is only one distinguishable way in which the atoms can be arranged. Consequently $S_1 = k \ln 1 = 0$ and therefore $\Delta S_{\text{mix}} = S_2$.

Assuming that A and B mix to form a substitutional solid solution and that all configurations of A and B atoms are equally probable, the number of distinguishable ways of arranging the atoms on the atom sites is

$$\omega_{\text{config}} = \frac{(N_A + N_B)!}{N_A! N_B!} \quad (1.24)$$

where N_A is the number of A atoms and N_B the number of B atoms.

Since we are dealing with 1 mol of solution, i.e. N_a atoms (Avogadro's number).

$$N_A = X_A N_a$$

and

$$N_B = X_B N_a$$

By substituting into Equations 1.23 and 1.24, using Stirling's approximation ($\ln N! \approx N \ln N - N$) and the relationship $N_a k = R$ (the universal gas constant) gives

$$\Delta S_{\text{mix}} = -R(X_A \ln X_A + X_B \ln X_B) \quad (1.25)$$

Note that, since X_A and X_B are less than unity, ΔS_{mix} is positive, i.e. there is an increase in entropy on mixing, as expected. The free energy of mixing, ΔG_{mix} , is obtained from Equation 1.22 as

$$\Delta G_{\text{mix}} = RT(X_A \ln X_A + X_B \ln X_B) \quad (1.26)$$

Figure 1.9 shows ΔG_{mix} as a function of composition and temperature.

The actual free energy of the solution G will also depend on G_A and G_B . From Equations 1.19, 1.20 and 1.26

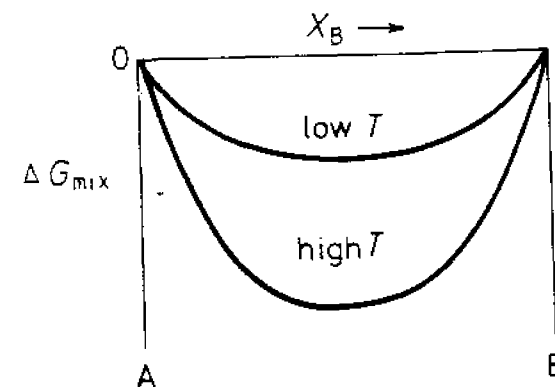


Fig. 1.9 Free energy of mixing for an ideal solution.

$$G = G_2 = X_A G_A + X_B G_B + RT(X_A \ln X_A + X_B \ln X_B) \quad (1.27)$$

This is shown schematically in Fig. 1.10. Note that, as the temperature increases, G_A and G_B decrease and the free energy curves assume a greater curvature. The decrease in G_A and G_B is due to the thermal entropy of both components and is given by Equation 1.10.

It should be noted that all of the free energy-composition diagrams in this book are essentially schematic; if properly plotted the free energy curves must end asymptotically at the vertical axes of the pure components, i.e. tangential to the vertical axes of the diagrams. This can be shown by differentiating Equation 1.26 or 1.27.

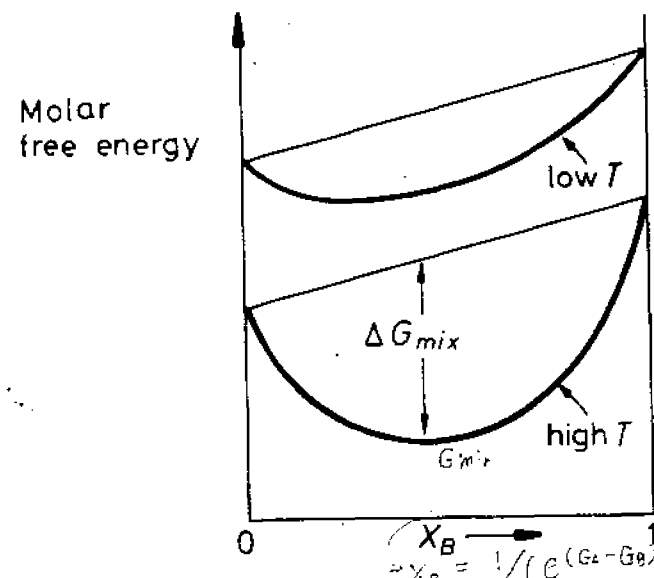


Fig. 1.10 The molar free energy (free energy per mole of solution) for an ideal solid solution. A combination of Figs. 1.8 and 1.9.

1.3.3 Chemical Potential

In alloys it is of interest to know how the free energy of a given phase will change when atoms are added or removed. If a small quantity of A, dn_A mol, is added to a large amount of a phase at constant temperature and pressure, the size of the system will increase by dn_A and therefore the *total* free energy of the system will also increase by a small amount dG' . If dn_A is small enough dG' will be proportional to the amount of A added. Thus we can write

$$dG' = \mu_A dn_A \quad (T, P, n_B \text{ constant}) \quad (1.28)$$

The proportionality constant μ_A is called the *partial molar free energy* of A or alternatively the *chemical potential* of A in the phase. μ_A depends on the composition of the phase, and therefore dn_A must be so small that the composition is not significantly altered. If Equation 1.28 is rewritten it can be seen that a definition of the chemical potential of A is

$$\mu_A = \left(\frac{\partial G'}{\partial n_A} \right)_{T, P, n_B} \quad (1.29)$$

The symbol G' has been used for the Gibbs free energy to emphasize the fact that it refers to the whole system. The usual symbol G will be used to denote the *molar* free energy and is therefore independent of the size of the system.

Equations similar to 1.28 and 1.29 can be written for the other components in the solution. For a binary solution at constant temperature and pressure the separate contributions can be summed:

$$dG' = \mu_A dn_A + \mu_B dn_B \quad (1.30)$$

This equation can be extended by adding further terms for solutions containing more than two components. If T and P changes are also allowed Equation 1.9 must be added giving the general equation

$$dG' = -SdT + VdP + \mu_A dn_A + \mu_B dn_B + \mu_C dn_C + \dots$$

If 1 mol of the original phase contained X_A mol A and X_B mol B, the size of the system can be increased without altering its composition if A and B are added in the correct proportions, i.e. such that $dn_A:dn_B = X_A:X_B$. For example if the phase contains twice as many A as B atoms ($X_A = 2/3$, $X_B = 1/3$) the composition can be maintained constant by adding two A atoms for every one B atom ($dn_A:dn_B = 2$). In this way the size of the system can be increased by 1 mol without changing μ_A and μ_B . To do this X_A mol A and X_B mol B must be added and the free energy of the system will increase by the molar free energy G . Therefore from Equation 1.30

$$G = \mu_A X_A + \mu_B X_B \quad \text{J mol}^{-1} \quad (1.31)$$

When G is known as a function of X_A and X_B , as in Fig. 1.10 for example, μ_A and μ_B can be obtained by extrapolating the tangent to the G curve to the

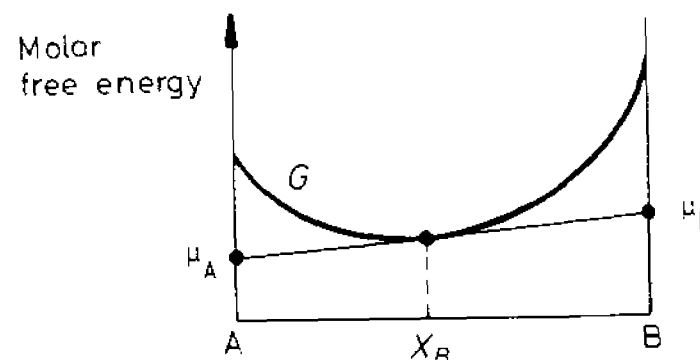


Fig. 1.11 The relationship between the free energy curve for a solution and the chemical potentials of the components.

sides of the molar free energy diagram as shown in Fig. 1.11. This can be obtained from Equations 1.30 and 1.31, remembering that $X_A + X_B = 1$, i.e. $dX_A = -dX_B$, and this is left as an exercise for the reader. It is clear from Fig. 1.11 that μ_A and μ_B vary systematically with the composition of the phase.

Comparison of Equations 1.27 and 1.31 gives μ_A and μ_B for an ideal solution as

$$\begin{aligned} \mu_A &= G_A + RT \ln X_A \\ \mu_B &= G_B + RT \ln X_B \end{aligned} \quad (1.32)$$

which is a much simpler way of presenting Equation 1.27. These relationships are shown in Fig. 1.12. The distances ac and bd are simply $-RT \ln X_A$ and $-RT \ln X_B$.

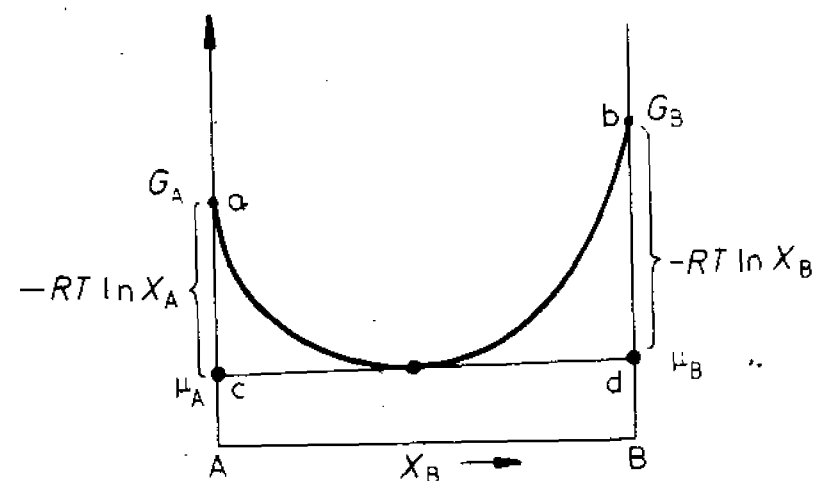


Fig. 1.12 The relationship between the free energy curve and chemical potentials for an ideal solution.

1.3.4 Regular Solutions

Returning to the model of a solid solution, so far it has been assumed that $\Delta H_{\text{mix}} = 0$; however, this type of behaviour is exceptional in practice and usually mixing is *endothermic* (heat absorbed) or *exothermic* (heat evolved). The simple model used for an ideal solution can, however, be extended to include the ΔH_{mix} term by using the so-called *quasi-chemical* approach.

In the quasi-chemical model it is assumed that the heat of mixing, ΔH_{mix} , is only due to the bond energies between adjacent atoms. For this assumption to be valid it is necessary that the volumes of pure A and B are equal and do not change during mixing so that the interatomic distances and bond energies are independent of composition.

The structure of a binary solid solution is shown schematically in Fig. 1.13. Three types of interatomic bonds are present:

1. A—A bonds each with an energy ϵ_{AA} .
2. B—B bonds each with an energy ϵ_{BB} .
3. A—B bonds each with an energy ϵ_{AB} .

By considering zero energy to be the state where the atoms are separated to infinity ϵ_{AA} , ϵ_{BB} and ϵ_{AB} are negative quantities, and become increasingly more negative as the bonds become stronger. The internal energy of the solution E will depend on the number of bonds of each type P_{AA} , P_{BB} and P_{AB} such that

$$E = P_{AA}\epsilon_{AA} + P_{BB}\epsilon_{BB} + P_{AB}\epsilon_{AB}$$

Before mixing pure A and B contain only A—A and B—B bonds respectively and by considering the relationships between P_{AA} , P_{BB} and P_{AB} in the solution it can be shown¹ that the change in internal energy on mixing is given

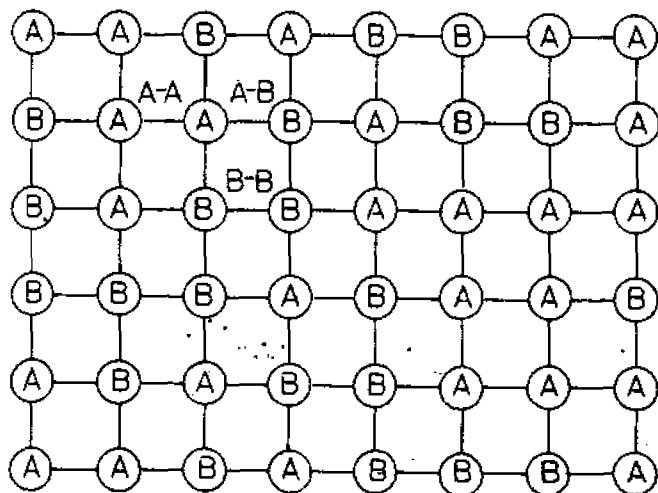


Fig. 1.13 The different types of interatomic bond in a solid solution.

by

$$\Delta H_{\text{mix}} = P_{AB}\epsilon \quad (1.33)$$

where

$$\epsilon = \epsilon_{AB} - \frac{1}{2}(\epsilon_{AA} + \epsilon_{BB}) \quad (1.34)$$

that is, ϵ is the difference between the A—B bond energy and the average of the A—A and B—B bond energies.

If $\epsilon = 0$, $\Delta H_{\text{mix}} = 0$ and the solution is ideal, as considered in Section 1.3.2. In this case the atoms are *completely randomly arranged* and the entropy of mixing is given by Equation 1.25. In such a solution it can also be shown¹ that

$$P_{AB} = N_a z X_A X_B \text{ bonds mol}^{-1} \quad (1.35)$$

where N_a is Avogadro's number, and z is the number of bonds per atom.

If $\epsilon < 0$ the atoms in the solution will prefer to be surrounded by atoms of the opposite type and this will increase P_{AB} , whereas, if $\epsilon > 0$, P_{AB} will tend to be less than in a random solution. However, provided ϵ is not too different from zero, Equation 1.35 is still a good approximation in which case

$$\Delta H_{\text{mix}} = \Omega X_A X_B \quad (1.36)$$

where

$$\Omega = N_a z \epsilon \quad (1.37)$$

Real solutions that closely obey Equation 1.36 are known as *regular solutions*. The variation of ΔH_{mix} with composition is parabolic and is shown in Fig. 1.14 for $\Omega > 0$. Note that the tangents at $X_A = 0$ and 1 are related to Ω as shown.

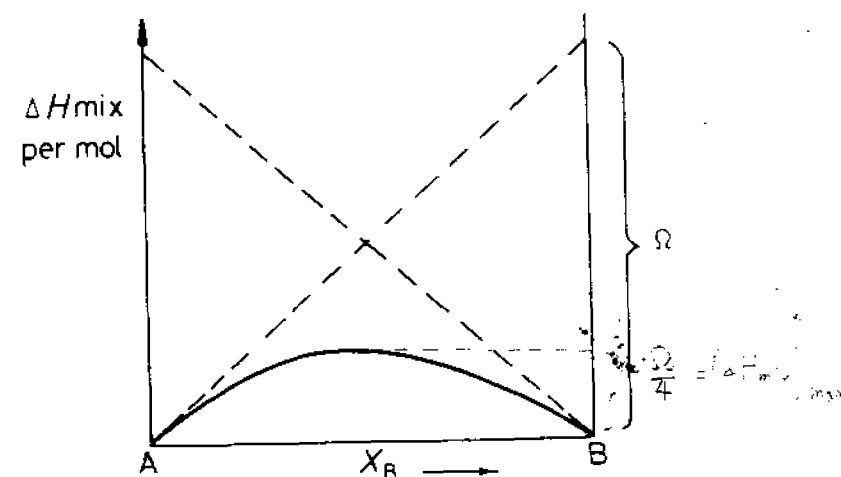


Fig. 1.14 The variation of ΔH_{mix} with composition for a regular solution.

The free energy change on mixing a regular solution is given by Equations 1.21, 1.25 and 1.36 as

$$\Delta G_{\text{mix}} = \frac{\Omega X_A X_B}{\Delta H_{\text{mix}}} + \underbrace{RT(X_A \ln X_A + X_B \ln X_B)}_{-T\Delta S_{\text{mix}}} \quad (1.38)$$

This is shown in Fig. 1.15 for different values of Ω and temperature. For exothermic solutions $\Delta H_{\text{mix}} < 0$ and mixing results in a free energy decrease at all temperatures (Fig. 1.15a and b). When $\Delta H_{\text{mix}} > 0$, however, the situation is more complicated. At high temperatures $T\Delta S_{\text{mix}}$ is greater than ΔH_{mix} for all compositions and the free energy curve has a positive curvature at all points (Fig. 1.15c). At low temperatures, on the other hand, $T\Delta S_{\text{mix}}$ is smaller and ΔG_{mix} develops a negative curvature in the middle (Fig. 1.15d).

Differentiating Equation 1.25 shows that, as X_A or $X_B \rightarrow 0$, the $-T\Delta S_{\text{mix}}$ curve becomes vertical whereas the slope of the ΔH_{mix} curve tends to a finite

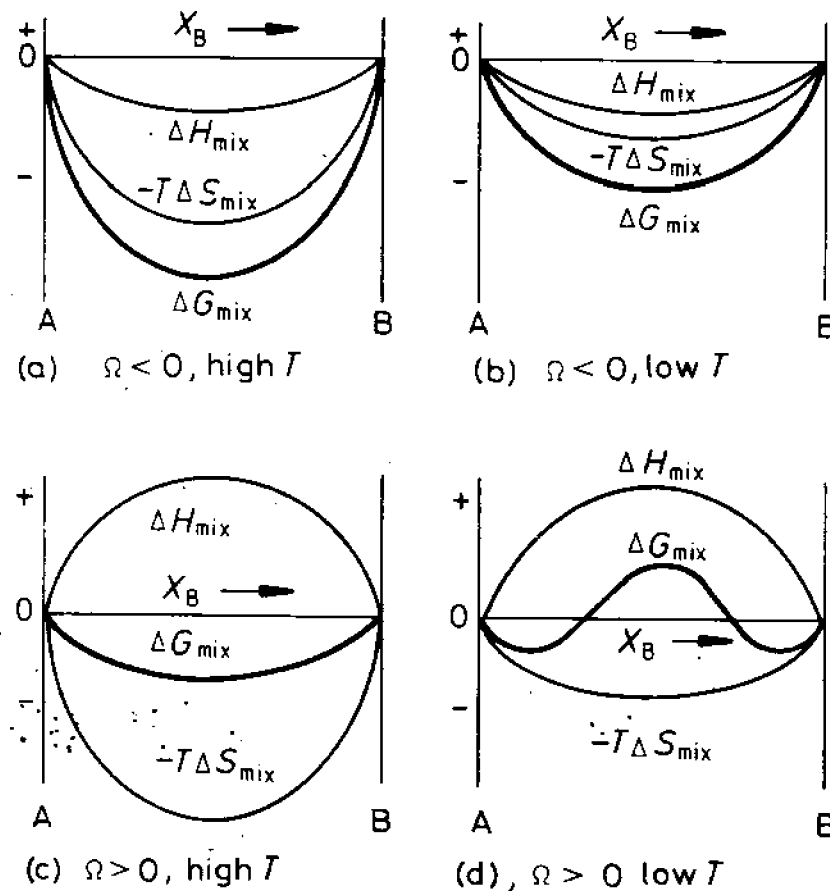


Fig. 1.15 The effect of ΔH_{mix} and T on ΔG_{mix} .

value Ω (Fig. 1.14). This means that, except at absolute zero, ΔG_{mix} always decreases on addition of a small amount of solute.

The actual free energy of the alloy depends on the values chosen for G_A and G_B and is given by Equations 1.19, 1.20 and 1.38 as

$$G = X_A G_A + X_B G_B + \Omega X_A X_B + RT(X_A \ln X_A + X_B \ln X_B) \quad (1.39)$$

This is shown in Fig. 1.16 along with the chemical potentials of A and B in the solution. Using the relationship $X_A X_B = X_A^2 X_B + X_B^2 X_A$ and comparing Equations 1.31 and 1.39 shows that for a regular solution

$$\mu_A = G_A + \Omega(1 - X_A)^2 + RT \ln X_A \quad (1.40)$$

and

$$\mu_B = G_B + \Omega(1 - X_B)^2 + RT \ln X_B$$

1.3.5 Activity

Expression 1.32 for the chemical potential of an ideal alloy was simple and it is convenient to retain a similar expression for any solution. This can be done by defining the *activity* of a component, a , such that the distances ac and bd in Fig. 1.16 are $-RT \ln a_A$ and $-RT \ln a_B$. In this case

$$\mu_A = G_A + RT \ln a_A \quad (1.41)$$

and

$$\mu_B = G_B + RT \ln a_B$$

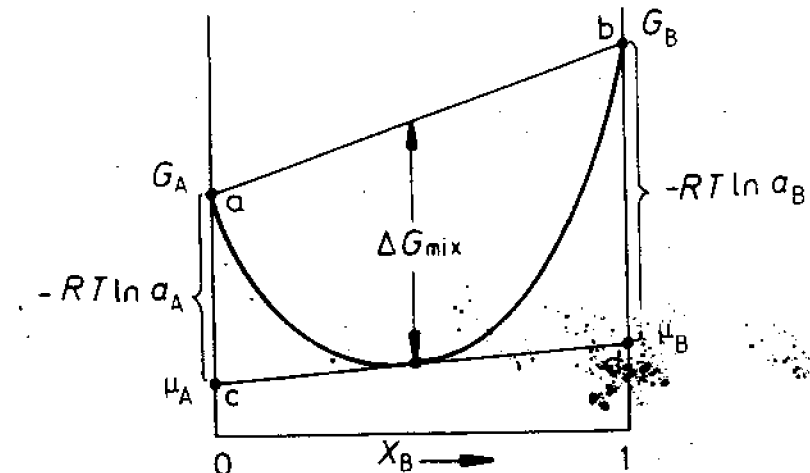


Fig. 1.16 The relationship between molar free energy and activity.

In general a_A and a_B will be different from X_A and X_B and the relationship between them will vary with the composition of the solution. For a regular solution, comparison of Equations 1.40 and 1.41 gives

$$\ln \left(\frac{a_A}{X_A} \right) = \frac{\Omega}{RT} (1 - X_A)^2$$

and

$$\ln \left(\frac{a_B}{X_B} \right) = \frac{\Omega}{RT} (1 - X_B)^2 \quad (1.42)$$

Assuming pure A and pure B have the same crystal structure, the relationship between a and X for any solution can be represented graphically as illustrated in Fig. 1.17. Line 1 represents an ideal solution for which $a_A = X_A$ and $a_B = X_B$. If $\Delta H_{\text{mix}} < 0$ the activity of the components in solution will be less in an ideal solution (line 2) and vice versa when $\Delta H_{\text{mix}} > 0$ (line 3).

The ratio (a_A/X_A) is usually referred to as γ_A , the activity coefficient of A, that is

$$\gamma_A = a_A/X_A \quad (1.43)$$

For a dilute solution of B in A, Equation 1.42 can be simplified by letting $X_B \rightarrow 0$ in which case

$$\gamma_B = \frac{a_B}{X_B} = \text{constant} \quad (\text{Henry's law}) \quad (1.44)$$

and

$$\gamma_A = \frac{a_A}{X_A} = 1 \quad (\text{Raoult's law}) \quad (1.45)$$

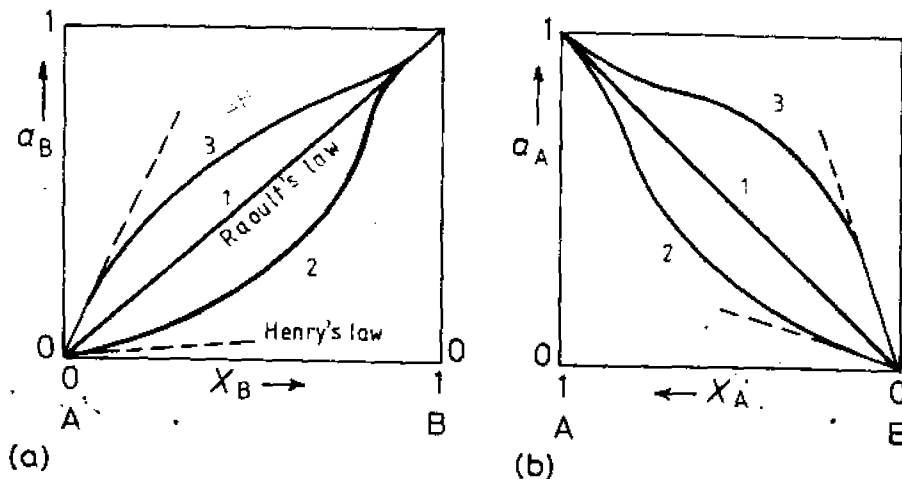


Fig. 1.17 The variation of activity with composition (a) a_B (b) a_A . Line 1: ideal solution (Raoult's law). Line 2: $\Delta H_{\text{mix}} < 0$. Line 3: $\Delta H_{\text{mix}} > 0$.

Equation 1.44 is known as Henry's law and 1.45 as Raoult's law; they apply to all solutions when sufficiently dilute.

Since activity is simply related to chemical potential via Equation 1.41 the activity of a component is just another means of describing the state of the component in a solution. No extra information is supplied and its use is simply a matter of convenience as it often leads to simpler mathematics.

Activity and chemical potential are simply a measure of the tendency of an atom to leave a solution. If the activity or chemical potential is low the atoms are reluctant to leave the solution which means, for example, that the vapour pressure of the component in equilibrium with the solution will be relatively low. It will also be apparent later that the activity or chemical potential of a component is important when several condensed phases are in equilibrium.

1.3.6 Real Solutions

While the previous model provides a useful description of the effects of configurational entropy and interatomic bonding on the free energy of binary solutions its practical use is rather limited. For many systems the model is an oversimplification of reality and does not predict the correct dependence of ΔG_{mix} on composition and temperature.

As already indicated, in alloys where the enthalpy of mixing is not zero (ϵ and $\Omega \neq 0$) the assumption that a random arrangement of atoms is the equilibrium, or most stable arrangement is not true, and the calculated value for ΔG_{mix} will not give the minimum free energy. The actual arrangement of atoms will be a compromise that gives the lowest internal energy consistent with sufficient entropy, or randomness, to achieve the minimum free energy. In systems with $\epsilon < 0$ the internal energy of the system is reduced by increasing the number of A—B bonds, i.e. by ordering the atoms as shown in Fig. 1.18a. If $\epsilon > 0$ the internal energy can be reduced by increasing the number of A—A and B—B bonds, i.e. by the clustering of the atoms into A-rich and B-rich groups, Fig. 1.18b. However, the degree of ordering or

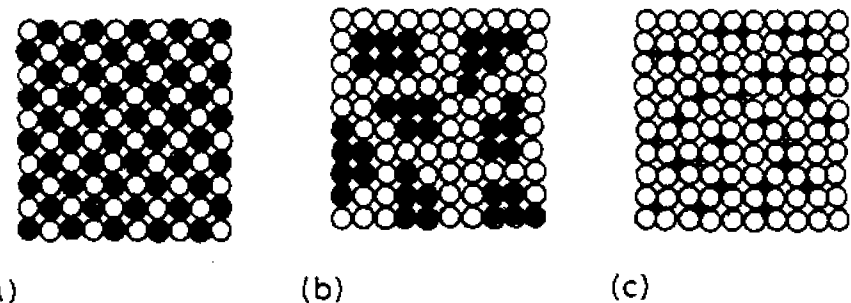


Fig. 1.18 Schematic representation of solid solutions: (a) ordered substitutional, (b) clustering, (c) random interstitial.

clustering will decrease as temperature increases due to the increasing importance of entropy.

In systems where there is a size difference between the atoms the quasi-chemical model will underestimate the change in internal energy on mixing since no account is taken of the elastic strain fields which introduce a strain energy term into ΔH_{mix} . When the size difference is large this effect can dominate over the chemical term.

When the size difference between the atoms is very large then *interstitial solid solutions* are energetically most favourable, Fig. 1.18c. New mathematical models are needed to describe these solutions.

In systems where there is strong chemical bonding between the atoms there is a tendency for the formation of intermetallic phases. These are distinct from solutions based on the pure components since they have a different crystal structure and may also be highly ordered. Intermediate phases and ordered phases are discussed further in the next two sections.

1.3.7 Ordered Phases

If the atoms in a substitutional solid solution are completely randomly arranged each atom position is equivalent and the probability that any given site in the lattice will contain an A atom will be equal to the fraction of A atoms in the solution X_A , similarly X_B for the B atoms. In such solutions P_{AB} , the number of A—B bonds, is given by Equation 1.35. If $\Omega < 0$ and the number of A—B bonds is greater than this, the solution is said to contain short-range order (SRO). The degree of ordering can be quantified by defining a SRO parameter s such that

$$s = \frac{P_{AB} - P_{AB}(\text{random})}{P_{AB}(\text{max}) - P_{AB}(\text{random})}$$

where $P_{AB}(\text{max})$ and $P_{AB}(\text{random})$ refer to the maximum number of bonds possible and the number of bonds for a random solution, respectively. Figure 1.19 illustrates the difference between random and short-range ordered solutions.

In solutions with compositions that are close to a simple ratio of A : B atoms another type of order can be found as shown schematically in Fig. 1.18a. This is known as long-range order. Now the atom sites are no longer equivalent but can be labelled as A-sites and B-sites. Such a solution can be considered to be a different (ordered) phase separate from the random or nearly random solution.

Consider Cu-Au alloys as a specific example. Cu and Au are both fcc and totally miscible. At high temperatures Cu or Au atoms can occupy any site and the lattice can be considered as fcc with a 'random' atom at each lattice

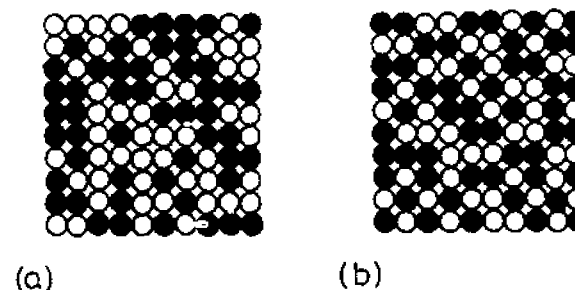


Fig. 1.19 (a) Random A—B solution with a total of 100 atoms and $X_A = X_B = 0.5$. $P_{AB} \sim 100$, $S = 0$. (b) Same alloy with short-range order $P_{AB} = 132$. $P_{AB(\text{max})} \sim 200$, $S = (132 - 100)/(200 - 100) = 0.32$.

point as shown in Fig. 1.20a. At low temperatures, however, solutions with $X_{\text{Cu}} = X_{\text{Au}} = 0.5$, i.e. a 50/50 Cu/Au mixture, form an ordered structure in which the Cu and Au atoms are arranged in alternate layers. Fig. 1.20b. Each atom position is no longer equivalent and the lattice is described as a CuAu superlattice. In alloys with the composition Cu_3Au another superlattice is found, Fig. 1.20c.

The entropy of mixing of structures with long-range order is extremely small and with increasing temperature the degree of order decreases until above some critical temperature there is no long-range order at all. This temperature is a maximum when the composition is the ideal required for the superlattice. However, long-range order can still be obtained when the composition deviates from the ideal if some of the atom sites are left vacant or if some atoms sit on wrong sites. In such cases it can be easier to disrupt the order with increasing temperature and the critical temperature is lower, see Fig. 1.21.

The most common ordered lattices in other systems are summarized in Fig. 1.22 along with their *Strukturbericht* notation and examples of alloys in which they are found. Finally, note that the critical temperature for loss of long-range order increases with increasing Ω , or ΔH_{mix} , and in many systems the ordered phase is stable up to the melting point.

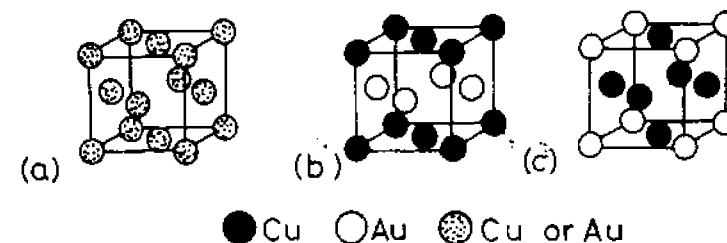


Fig. 1.20 Ordered substitutional structures in the Cu—Au system: (a) high-temperature disordered structure, (b) CuAu superlattice, (c) Cu_3Au superlattice.

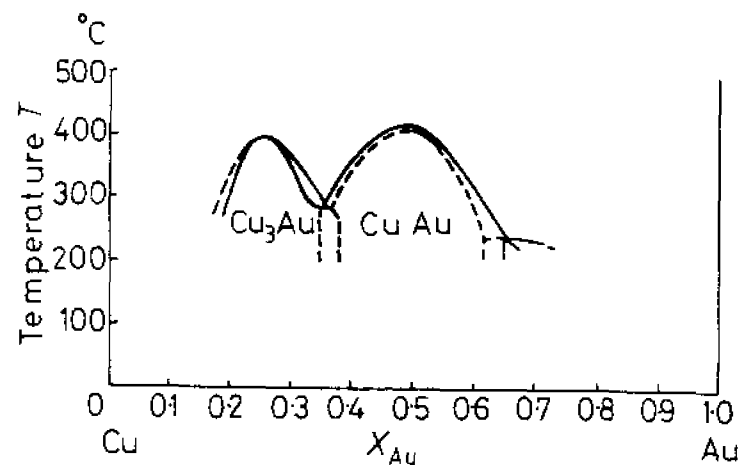


Fig. 1.21 Part of the Cu-Au phase diagram showing the regions where the Cu_3Au and CuAu superlattices are stable.

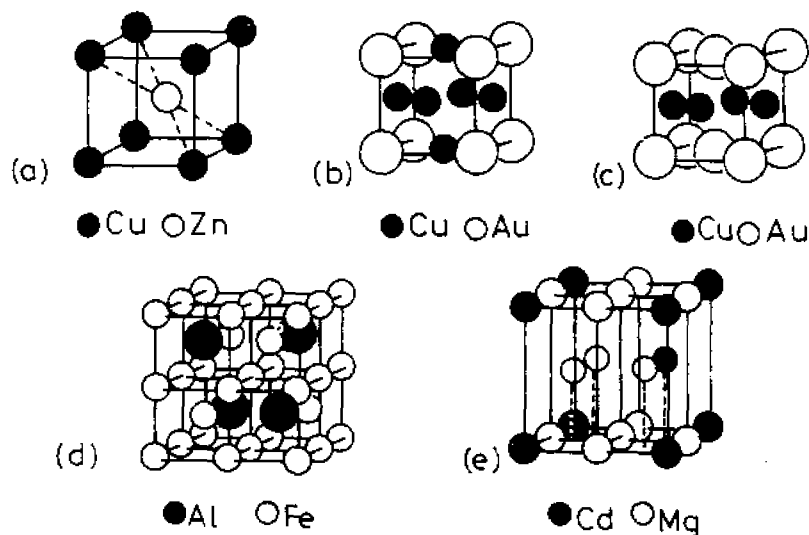


Fig. 1.22 The five common ordered lattices, examples of which are: (a) L_{20} : CuZn , FeCo , NiAl , FeAl , AgMg ; (b) L_{12} : Cu_3Au , Au_3Cu , Ni_3Mn , Ni_3Fe , Ni_3Al , Pt_3Fe ; (c) L_{10} : CuAu , CoPt , FePt ; (d) D_{03} : Fe_3Al , Fe_3Si , Fe_3Be , Cu_3Al ; (e) D_{019} : Mg_3Cd , Cd_3Mg , Ti_3Al , Ni_3Sn . (After R.E. Smallman, *Modern Physical Metallurgy*, 3rd edition, Butterworths, London, 1970.)

1.3.8 Intermediate Phases

Often the configuration of atoms that has the minimum free energy after mixing does not have the same crystal structure as either of the pure components. In such cases the new structure is known as an *intermediate phase*.

Intermediate phases are often based on an ideal atom ratio that results in a minimum Gibbs free energy. For compositions that deviate from the ideal, the free energy is higher giving a characteristic 'U' shape to the G curve, as in Fig. 1.23. The range of compositions over which the free energy curve has a meaningful existence depends on the structure of the phase and the type of interatomic bonding—metallic, covalent or ionic. When small composition deviations cause a rapid rise in G the phase is referred to as an *intermetallic compound* and is usually *stoichiometric*, i.e. has a formula A_mB_n where m and n are integers, Fig. 1.23a. In other structures fluctuations in composition can be tolerated by some atoms occupying 'wrong' positions or by atom sites being left vacant, and in these cases the curvature of the G curve is much less, Fig. 1.23b.

Some intermediate phases can undergo order-disorder transformations in which an almost random arrangement of the atoms is stable at high temperatures and an ordered structure is stable below some critical temperature. Such a transformation occurs in the β phase in the Cu-Zn system for example (see Section 5.10).

The structure of intermediate phases is determined by three main factors: relative atomic size, valency and electronegativity. When the component atoms differ in size by a factor of about 1.1–1.6 it is possible for the atoms to fill space most efficiently if the atoms order themselves into one of the so-called Laves phases based on MgCu_2 , MgZn_2 and MgNi_2 , Fig 1.24. Another example where atomic size determines the structure is in the formation of the *interstitial compounds* MX , M_2X , MX_2 and M_6X where M can be Zr, Ti, V, Cr, etc. and X can be H, B, C and N. In this case the M atoms form a cubic or hexagonal close-packed arrangement and the X atoms are small enough to fit into the interstices between them.

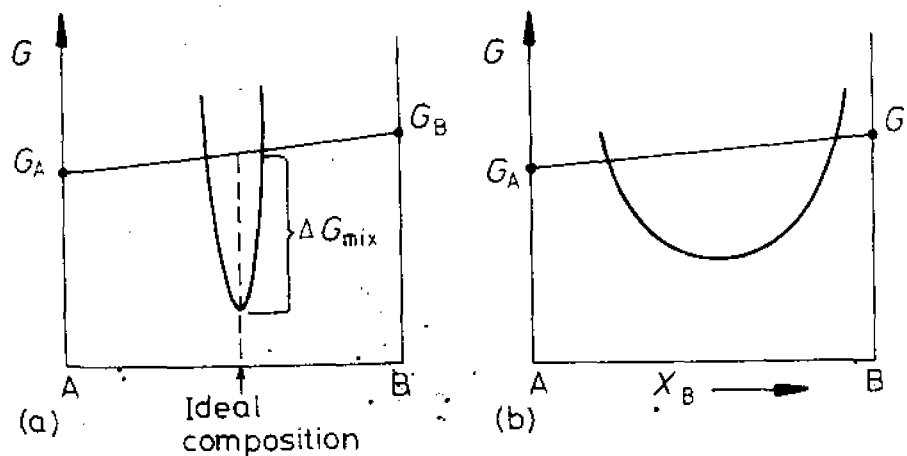


Fig. 1.23 Free energy curves for intermediate phases: (a) for an intermetallic compound with a very narrow stability range. (b) for an intermediate phase with a wide stability range.

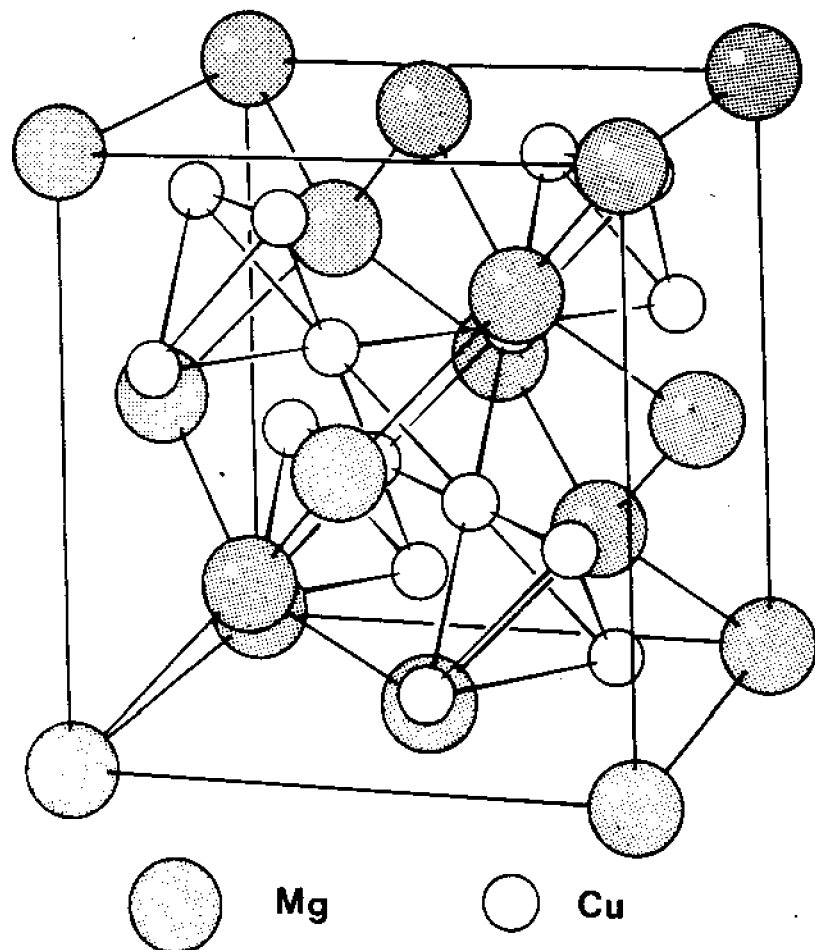


Fig. 1.24 The structure of MgCu_2 (A Laves phase). (From J.H. Wernick, chapter 5 in *Physical Metallurgy*, 2nd edn., R.W. Cahn (Ed.) North Holland, 1974.)

The relative valency of the atoms becomes important in the so-called *electron phases*, e.g. α and β brasses. The free energy of these phases depends on the number of valency electrons per unit cell, and this varies with composition due to the valency difference.

The electronegativity of an atom is a measure of how strongly it attracts electrons and in systems where the two components have very different electronegativities ionic bonds can be formed producing normal valency compounds, e.g. Mg^{2+} and Sn^{4-} are ionically bonded in Mg_2Sn .²

1.4 Equilibrium in Heterogeneous Systems

It is usually the case that A and B do not have the same crystal structure in their pure states at a given temperature. In such cases two free energy curves

must be drawn, one for each structure. The stable forms of pure A and B at a given temperature (and pressure) can be denoted as α and β respectively. For the sake of illustration let α be fcc and β bcc. The molar free energies of fcc A and bcc B are shown in Fig. 1.25a as points a and b. The first step in drawing the free energy curve of the fcc α phase is, therefore, to convert the stable bcc arrangement of B atoms into an unstable fcc arrangement. This requires an increase in free energy, bc. The free energy curve for the α phase can now be constructed as before by mixing fcc A and fcc B as shown in the figure. $-\Delta G_{\text{mix}}$ for α of composition X is given by the distance de as usual.

A similar procedure produces the molar free energy curve for the β phase. Fig. 1.25b. The distance af is now the difference in free energy between bcc A and fcc A.

It is clear from Fig. 1.25b that A-rich alloys will have the lowest free energy as a homogeneous α phase and B-rich alloys as β phase. For alloys with

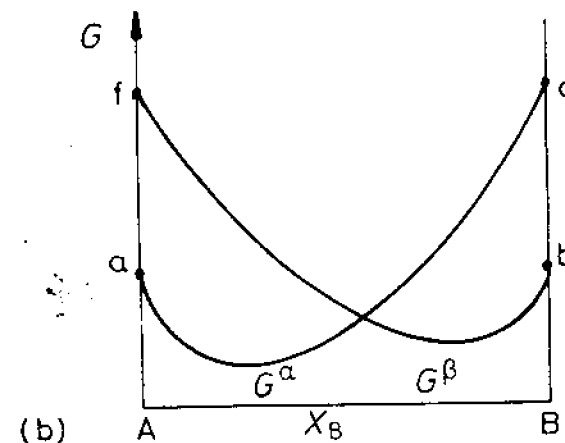
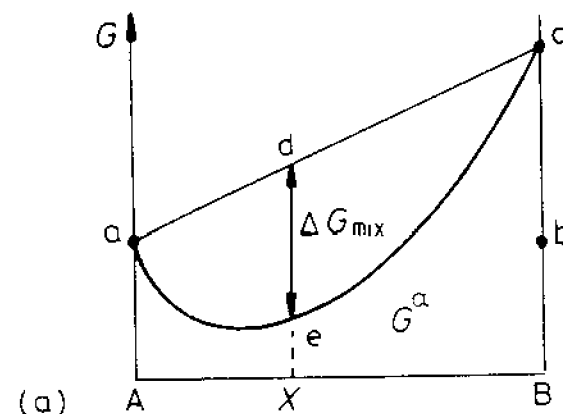


Fig. 1.25 (a) The molar free energy curve for the α phase. (b) Molar free energy curves for α and β phases.

compositions near the cross-over in the G curves the situation is not so straightforward. In this case it can be shown that the total free energy can be minimized by the atoms separating into two phases.

It is first necessary to consider a general property of molar free energy diagrams when phase mixtures are present. Suppose an alloy consists of two phases α and β each of which has a molar free energy given by G^α and G^β , Fig. 1.26. If the overall composition of the phase mixture is X_B^0 the lever rule gives the relative number of moles of α and β that must be present, and the molar free energy of the phase mixture G is given by the point on the straight line between α and β as shown in the figure. This result can be proven most readily using the geometry of Fig. 1.26. The lengths ad and cf respectively represent the molar free energies of the α and β phases present in the alloy. Point g is obtained by the intersection of be and dc so that bcg and acd , as well as deg and dfc , form similar triangles. Therefore $bg/ad = bc/ac$ and $ge/cf = ab/ac$. According to the lever rule 1 mol of alloy will contain bc/ac mol of α and ab/ac mol of β . It follows that bg and ge represent the separate contributions from the α and β phases to the total free energy of 1 mol of alloy. Therefore the length 'be' represents the molar free energy of the phase mixture.

Consider now alloy X^0 in Fig. 1.27a. If the atoms are arranged as a homogeneous phase, the free energy will be lowest as α , i.e. G_0^α per mole. However, from the above it is clear that the system can lower its free energy if the atoms separate into two phases with compositions α_1 and β_1 for example. The free energy of the system will then be reduced to G_1 . Further reductions in free energy can be achieved if the A and B atoms interchange between the α and β phases until the compositions α_e and β_e are reached, Fig. 1.27b. The free energy of the system G_e is now a minimum and there is no desire for further change. Consequently the system is in equilibrium and α_e and β_e are the equilibrium compositions of the α and β phases.

This result is quite general and applies to any alloy with an overall composition between α_e and β_e : only the relative amounts of the two phases change,

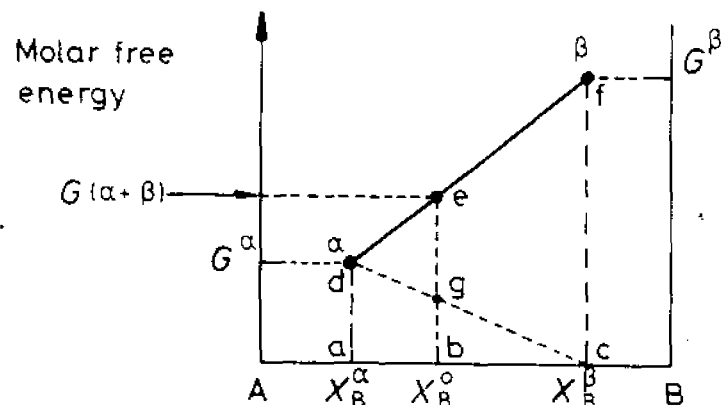


Fig. 1.26 The molar free energy of a two-phase mixture ($\alpha - \beta$).

as given by the lever rule. When the alloy composition lies outside this range, however, the minimum free energy lies on the G^α or G^β curves and the equilibrium state of the alloy is a homogeneous single phase.

From Fig. 1.27 it can be seen that equilibrium between two phases requires that the tangents to each G curve at the equilibrium compositions lie on a common line. In other words each component must have the same chemical potential in the two phases, i.e. for heterogeneous equilibrium:

$$\mu_A^\alpha = \mu_A^\beta, \quad \mu_B^\alpha = \mu_B^\beta \quad (1.46)$$

The condition for equilibrium in a heterogeneous system containing

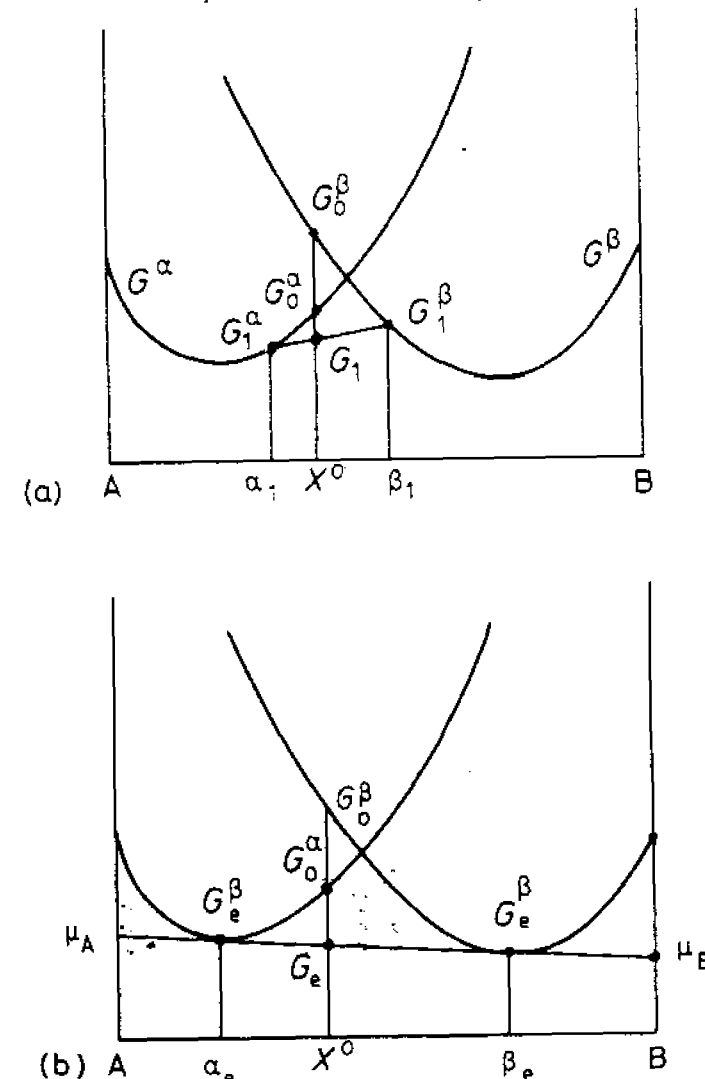


Fig. 1.27 (a) Alloy X^0 has a free energy G_1 as a mixture of $\alpha_1 + \beta_1$. (b) At equilibrium, alloy X^0 has a minimum free energy G_e when it is a mixture of $\alpha_e + \beta_e$.

two phases can also be expressed using the activity concept defined for homogeneous systems in Fig. 1.16. In heterogeneous systems containing more than one phase the pure components can, at least theoretically, exist in different crystal structures. The most stable state, with the lowest free energy, is usually defined as the state in which the pure component has unit activity. In the present example this would correspond to defining the activity of A in pure $\alpha - A$ as unity, i.e. when $X_A = 1$, $a_A^\alpha = 1$. Similarly when $X_B = 1$, $a_B^\beta = 1$. This definition of activity is shown graphically in Fig. 1.28a; Fig. 1.28b and c show how the activities of B and A vary with the composition of the α and β phases. Between A and α_e , and β_e and B, where single phases are stable, the activities (or chemical potentials) vary and for simplicity ideal solutions have been assumed in which case there is a straight line relationship between a and X . Between α_e and β_e the phase compositions in equilibrium do not change and the activities are equal and given by points q and r. In other words, when two phases exist in equilibrium, the activities of the components in the system must be equal in the two phases, i.e.

$$a_A^\alpha = a_A^\beta, \quad a_B^\alpha = a_B^\beta \quad (1.47)$$

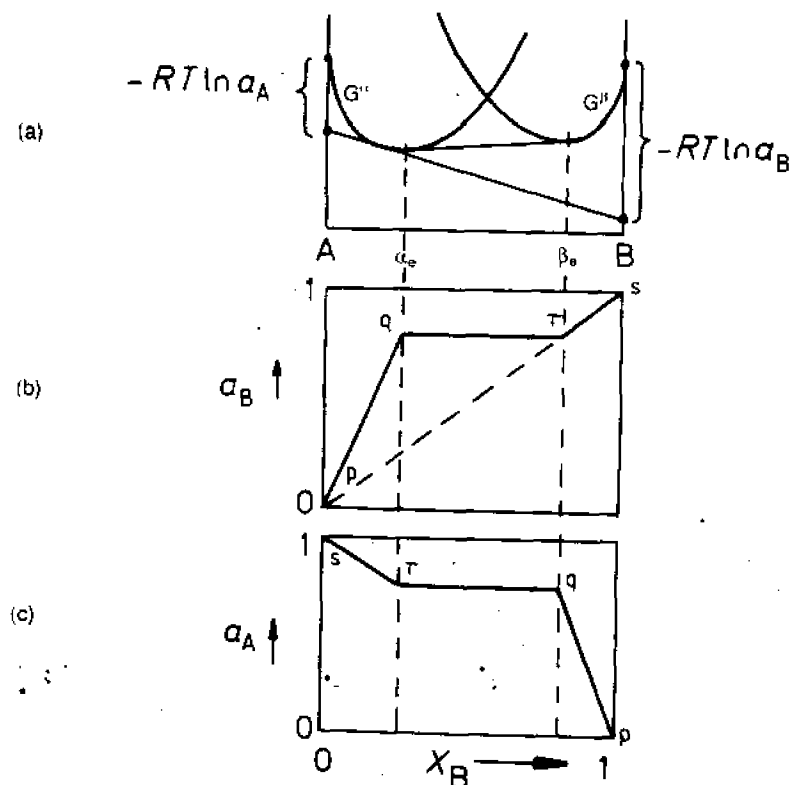


Fig. 1.28 The variation of a_A and a_B with composition for a binary system containing two ideal solutions, α and β .

1.5 Binary Phase Diagrams

In the previous section it has been shown how the equilibrium state of an alloy can be obtained from the free energy curves at a given temperature. The next step is to see how equilibrium is affected by temperature.

1.5.1 A Simple Phase Diagram

The simplest case to start with is when A and B are completely miscible in both the solid and liquid states and both are ideal solutions. The free energy of pure A and pure B will vary with temperature as shown schematically in Fig. 1.4. The equilibrium melting temperatures of the pure components occur when $G^S = G^L$, i.e. at $T_m(A)$ and $T_m(B)$. The free energy of both phases decreases as temperature increases. These variations are important for A-B alloys also since they determine the relative positions of G_A^S , G_A^L , G_B^S and G_B^L on the molar free energy diagrams of the alloy at different temperatures, Fig. 1.29.

At a high temperature $T_1 > T_m(A) > T_m(B)$ the liquid will be the stable phase for pure A and pure B, and for the simple case we are considering the liquid also has a lower free energy than the solid at all the intermediate compositions as shown in Fig. 1.29a.

Decreasing the temperature will have two effects: firstly G_A^L and G_B^L will increase more rapidly than G_A^S and G_B^S , secondly the curvature of the G curves will be reduced due to the smaller contribution of $-T\Delta S_{mix}$ to the free energy.

At $T_m(A)$, Fig. 1.29b, $G_A^S = G_A^L$, and this corresponds to point a on the A-B phase diagram, Fig. 1.29f. At a lower temperature T_2 the free energy curves cross, Fig. 1.29c, and the common tangent construction indicates that alloys between A and b are solid at equilibrium, between c and B they are liquid, and between b and c equilibrium consists of a two-phase mixture (S + L) with compositions b and c. These points are plotted on the equilibrium phase diagram at T_2 .

Between T_2 and $T_m(B)$ G^L continues to rise faster than G^S so that points b and c in Fig. 1.29c will both move to the right tracing out the solidus and liquidus lines in the phase diagram. Eventually at $T_m(B)$ b and c will meet at a single point, d in Fig. 1.29f. Below $T_m(B)$ the free energy of the solid phase is everywhere below that of the liquid and all alloys are stable as a single phase solid.

1.5.2 Systems with a Miscibility Gap

Figure 1.30 shows the free energy curves for a system in which the liquid phase is approximately ideal, but for the solid phase $\Delta H_{mix} > 0$, i.e. the A and B atoms 'dislike' each other. Therefore at low temperatures (T_3) the free energy curve for the solid assumes a negative curvature in the middle,

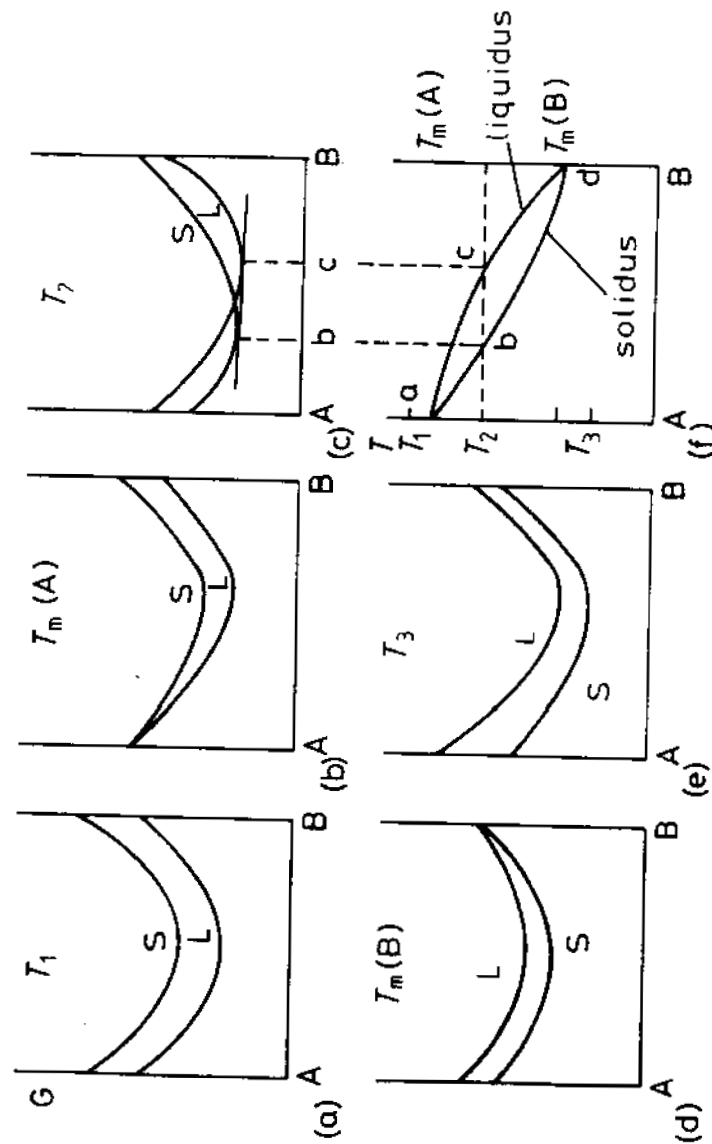


Fig. 1.29 The derivation of a simple phase diagram from the free energy curves for the liquid (L) and solid (S).

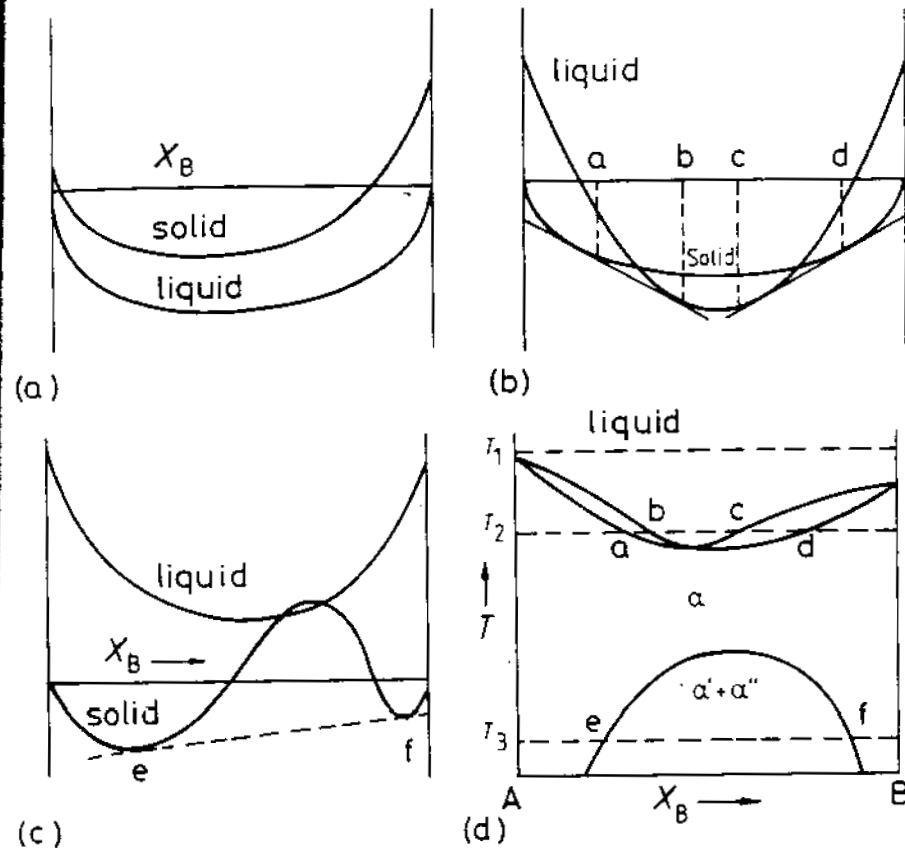


Fig. 1.30 The derivation of a phase diagram where $\Delta H_{\text{mix}}^{\text{S}} > \Delta H_{\text{mix}}^{\text{L}} = 0$. Free energy $v.$ composition curves for (a) T_1 , (b) T_2 , and (c) T_3 .

Fig. 1.30c, and the solid solution is most stable as a mixture of two phases α' and α'' with compositions e and f. At higher temperatures, when $-T\Delta S_{\text{mix}}$ becomes larger, e and f approach each other and eventually disappear as shown in the phase diagram, Fig. 1.30d. The $\alpha' + \alpha''$ region is known as a miscibility gap.

The effect of a positive ΔH_{mix} in the solid is already apparent at higher temperatures where it gives rise to a minimum melting point mixture. The reason why all alloys should melt at temperatures below the melting points of both components can be qualitatively understood since the atoms in the alloy 'repel' each other making the disruption of the solid into a liquid phase possible at lower temperatures than in either pure A or pure B.

1.5.3 Ordered Alloys

The opposite type of effect arises when $\Delta H_{\text{mix}} < 0$. In these systems melting will be more difficult in the alloys and a maximum melting point mixture may

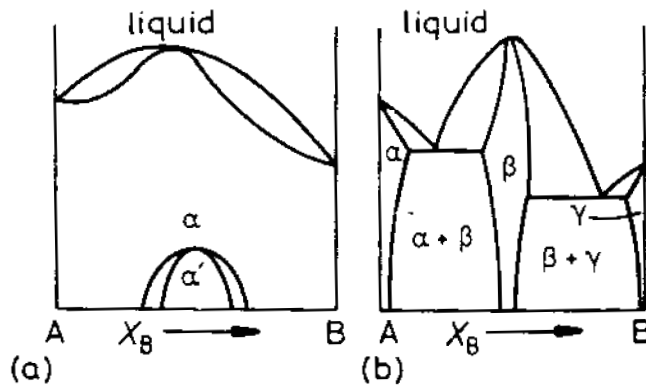


Fig. 1.31 (a) Phase diagram when $\Delta H_{\text{mix}}^S < 0$; (b) as (a) but even more negative ΔH_{mix}^S . (After R.A. Swalin, *Thermodynamics of Solids*, John Wiley, New York, 1972).

appear. This type of alloy also has a tendency to order at low temperatures as shown in Fig. 1.31a. If the attraction between unlike atoms is very strong the ordered phase may extend as far as the liquid, Fig. 1.31b.

1.5.4 Simple Eutectic Systems

If ΔH_{mix}^S is much larger than zero the miscibility gap in Fig. 1.30d can extend into the liquid phase. In this case a simple eutectic phase diagram results as shown in Fig. 1.32. A similar phase diagram can result when A and B have different crystal structures as illustrated in Fig. 1.33

1.5.5 Phase Diagrams Containing Intermediate Phases

When stable intermediate phases can form, extra free energy curves appear in the phase diagram. An example is shown in Fig. 1.34, which also illustrates how a peritectic transformation is related to the free energy curves.

An interesting result of the common tangent construction is that the stable composition range of the phase in the phase diagram need not include the composition with the minimum free energy, but is determined by the relative free energies of adjacent phases, Fig. 1.35. This can explain why the composition of the equilibrium phase appears to deviate from that which would be predicted from the crystal structure. For example the θ phase in the Cu-Al system is usually denoted as CuAl_2 although the composition $X_{\text{Cu}} = 1/3$, $X_{\text{Al}} = 2/3$ is not covered by the θ field on the phase diagram.

1.5.6 The Gibbs Phase Rule

The condition for equilibrium in a binary system containing two phases is given by Equation 1.46 or 1.47. A more general requirement for systems

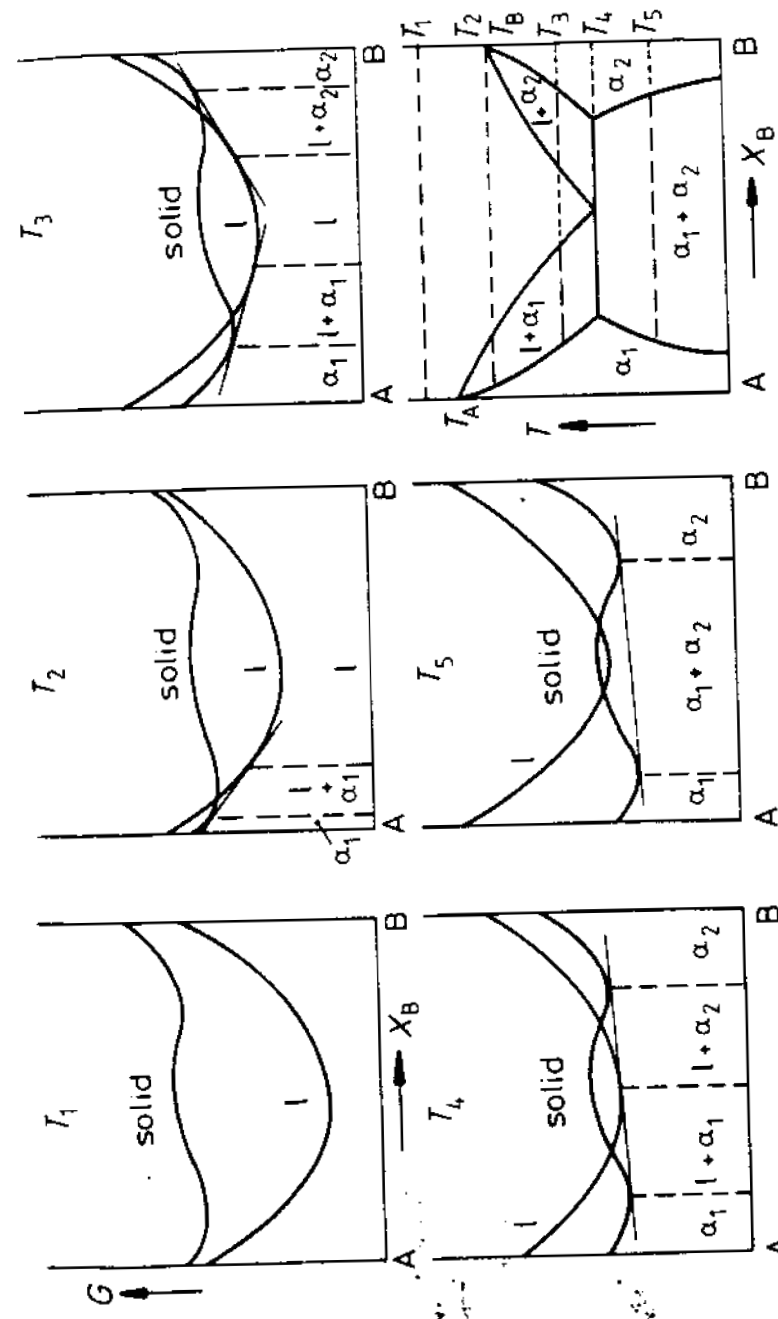


Fig. 1.32 The derivation of a eutectic phase diagram where both solid phases have the same crystal structure. (After A.H. Cottrell, *Theoretical Structural Metallurgy*, Edward Arnold, London, 1955, © Sir Alan Cottrell.)

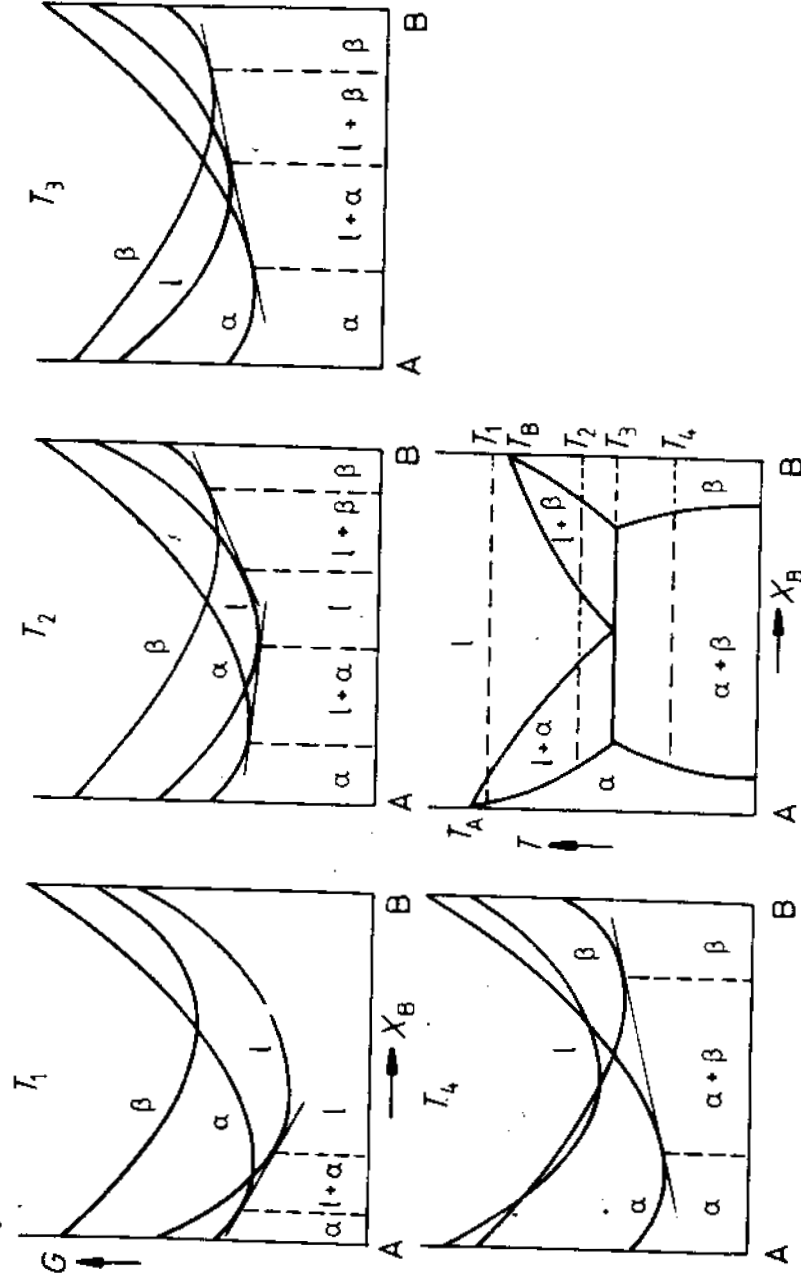


Fig. 1.33 The derivation of a eutectic phase diagram where each solid phase has a different crystal structure. (After A. Prince, *Alloy Phase Equilibria*, Elsevier, Amsterdam, 1966.)

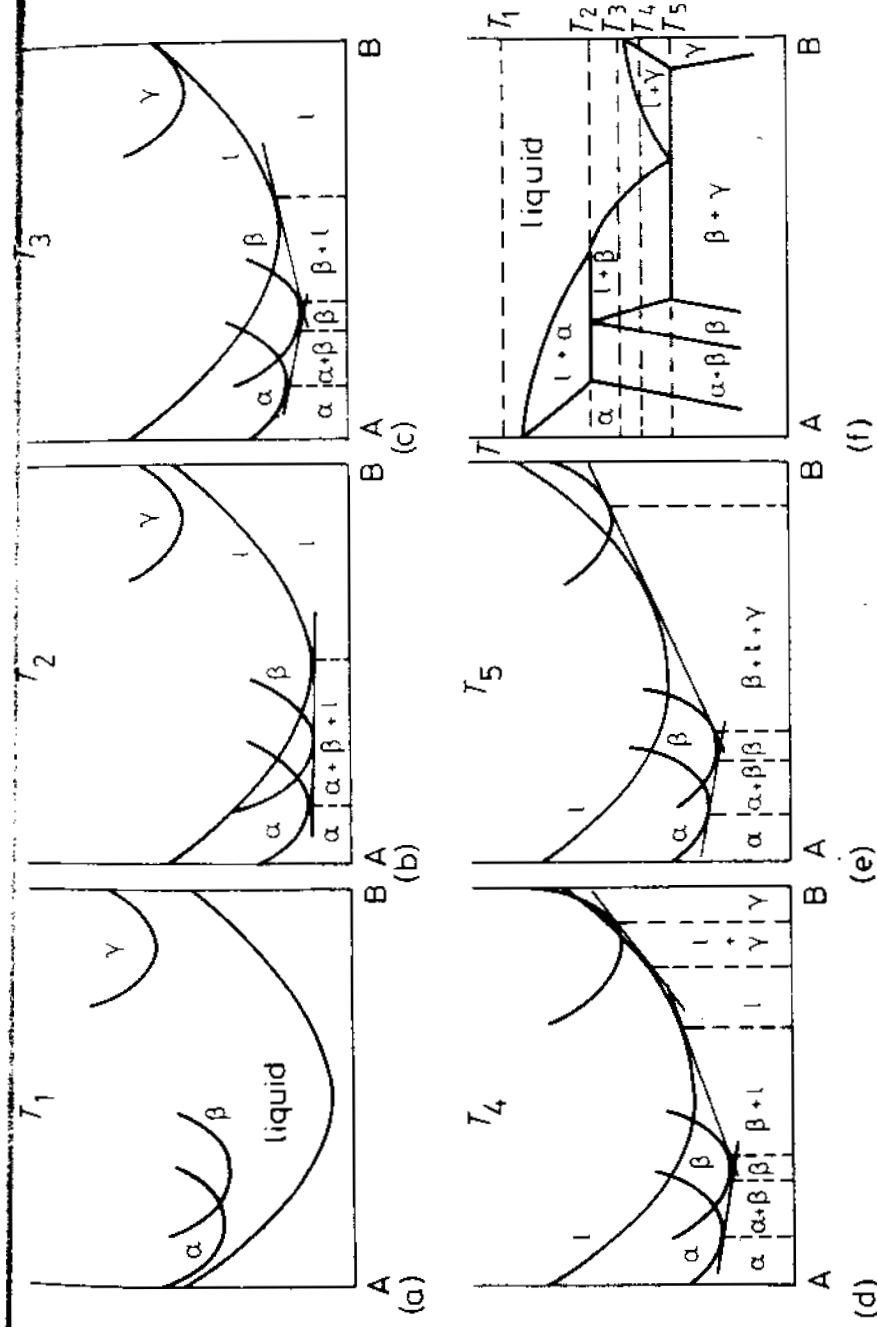


Fig. 1.34 The derivation of a complex phase diagram. (After A.H. Cottrell, *Theoretical Structural Metallurgy*, Edward Arnold, London, 1955, © Sir Alan Cottrell.)

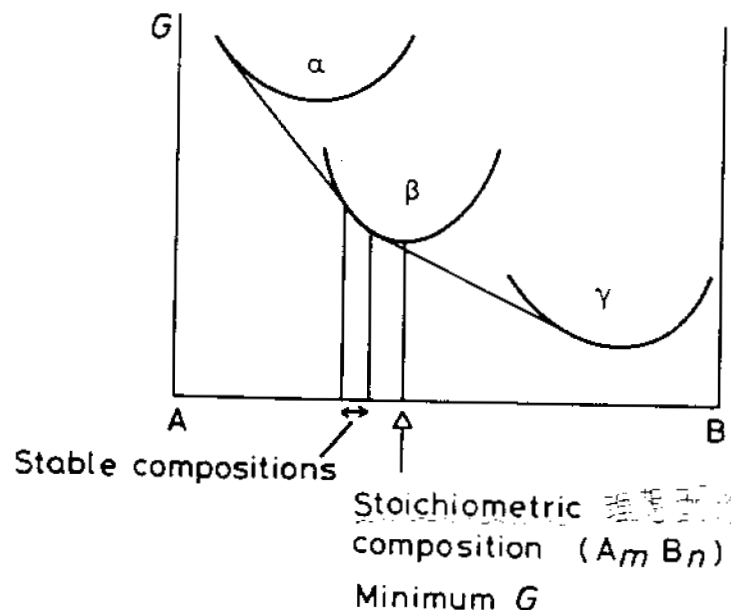


Fig. 1.35 Free energy diagram to illustrate that the range of compositions over which a phase is stable depends on the free energies of the other phases in equilibrium.

containing several components and phases is that the chemical potential of each component must be identical in every phase, i.e.

$$\begin{aligned}\mu_A^\alpha &= \mu_A^\beta = \mu_A^\gamma = \dots \\ \mu_B^\alpha &= \mu_B^\beta = \mu_B^\gamma = \dots \\ \mu_C^\alpha &= \mu_C^\beta = \mu_C^\gamma = \dots\end{aligned}\quad (1.48)$$

The proof of this relationship is left as an exercise for the reader (see Exercise 1.10). A consequence of this general condition is the Gibbs phase rule. This states that if a system containing C components and P phases is in equilibrium the number of degrees of freedom F is given by

$$P + F = C + 2 \quad (1.49)$$

A degree of freedom is an intensive variable such as T , P , X_A , X_B ... that can be varied independently while still maintaining equilibrium. If pressure is maintained constant one degree of freedom is lost and the phase rule becomes

$$P + F = C + 1 \quad (1.50)$$

At present we are considering binary alloys so that $C = 2$ therefore

$$P + F = 3$$

This means that a binary system containing one phase has two degrees of freedom, i.e. T and X_B can be varied independently. In a two-phase region of

a phase diagram $P = 2$ and therefore $F = 1$ which means that if the temperature is chosen independently the compositions of the phases are fixed. When three phases are in equilibrium, such as at a eutectic or peritectic temperature, there are no degrees of freedom and the compositions of the phases and the temperature of the system are all fixed.

1.5.7 The Effect of Temperature on Solid Solubility

The equations for free energy and chemical potential can be used to derive the effect of temperature on the limits of solid solubility in a terminal solid solution. Consider for simplicity the phase diagram shown in Fig. 1.36a where B is soluble in A, but A is virtually insoluble in B. The corresponding free energy curves for temperature T_1 are shown schematically in Fig. 1.36b. Since A is almost insoluble in B the G^β curve rises rapidly as shown. Therefore the maximum concentration of B soluble in A (X_B^c) is given by the condition

$$\mu_B^\alpha = \mu_B^\beta = G_B^\beta$$

For a regular solid solution Equation 1.40 gives

$$\mu_B^\alpha = G_B^\alpha + \Omega(1 - X_B)^2 + RT \ln X_B$$

But from Fig. 1.36b, $G_B^\alpha - \mu_B^\alpha = \Delta G_B$, the difference in free energy between pure B in the stable β -form and the unstable α -form. Therefore for $X_B = X_B^c$

$$-RT \ln X_B^c - \Omega(1 - X_B^c)^2 = \Delta G_B \quad (1.51)$$

If the solubility is low $X_B^c \ll 1$ and this gives

$$X_B^c = \exp \left\{ -\frac{\Delta G_B + \Omega}{RT} \right\} \quad (1.52)$$

Putting

$$\Delta G_B = \Delta H_B - T\Delta S_B$$

gives

$$X_B^c = A \exp \frac{-Q}{RT} \quad (1.53)$$

where A is a constant equal to $\exp(\Delta S_B/R)$ and

$$Q = \Delta H_B + \Omega > 0 \quad (1.54)$$

ΔH_B is the difference in enthalpy between the β -form of B and the α -form in J mol^{-1} . Ω is the change in energy when 1 mol of B with the α -structure dissolves in A to make a dilute solution. Therefore Q is just the enthalpy change, or heat absorbed, when 1 mol of B with the β -structure dissolves in A to make a dilute solution.

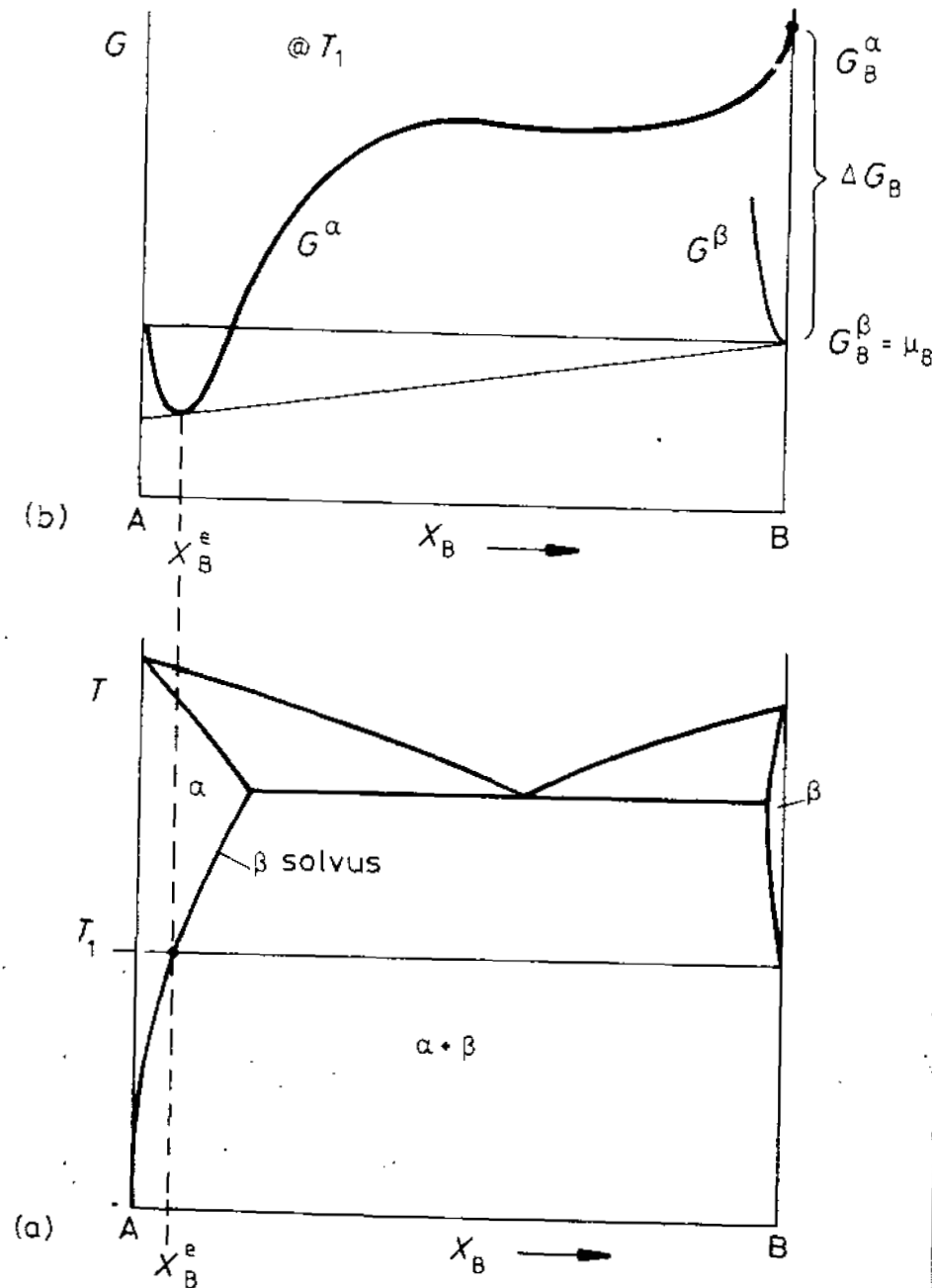


Fig. 1.36 Solubility of B in A.

ΔS_B is the difference in entropy between β -B and α -B and is approximately independent of temperature. Therefore the solubility of B in α increases exponentially with temperature at a rate determined by Q . It is interesting to note that, except at absolute zero, X_B^e can never be equal to zero, that is, no two components are ever completely insoluble in each other.

1.5.8 Equilibrium Vacancy Concentration

So far it has been assumed that in a metal lattice every atom site is occupied. However, let us now consider the possibility that some sites remain without atoms, that is, there are *vacancies* in the lattice. The removal of atoms from their sites not only increases the internal energy of the metal, due to the broken bonds around the vacancy, but also increases the randomness or configurational entropy of the system. The free energy of the alloy will depend on the concentration of vacancies and the equilibrium concentration X_v^e will be that which gives the minimum free energy.

If, for simplicity, we consider vacancies in a pure metal the problem of calculating X_v^e is almost identical to the calculation of ΔG_{mix} for A and B atoms when ΔH_{mix} is positive. Because the equilibrium concentration of vacancies is small the problem is simplified because vacancy-vacancy interactions can be ignored and the increase in enthalpy of the solid (ΔH) is directly proportional to the number of vacancies added, i.e.

$$\Delta H \approx \Delta H_v X_v$$

where X_v is the mole fraction of vacancies and ΔH_v is the increase in enthalpy per mole of vacancies added. (Each vacancy causes an increase of $\Delta H_v/N_a$ where N_a is Avogadro's number.)

There are two contributions to the entropy change ΔS on adding vacancies. There is a small change in the thermal entropy of ΔS_v per mole of vacancies added due to changes in the vibrational frequencies of the atoms around a vacancy. The largest contribution, however, is due to the increase in configurational entropy given by Equation 1.25. The total entropy change is thus

$$\Delta S = X_v \Delta S_v - R(X_v \ln X_v + (1 - X_v) \ln (1 - X_v))$$

The molar free energy of the crystal containing X_v mol of vacancies is therefore given by

$$G = G_A + \Delta G = G_A + \Delta H_v X_v - T \Delta S_v X_v + RT(X_v \ln X_v + (1 - X_v) \ln (1 - X_v)) \quad (1.55)$$

This is shown schematically in Fig. 1.37. Given time the number of vacancies will adjust so as to reduce G to a minimum. The equilibrium concentration of vacancies X_v^e is therefore given by the condition

$$\left. \frac{dG}{dX_v} \right|_{X_v = X_v^e} = 0$$

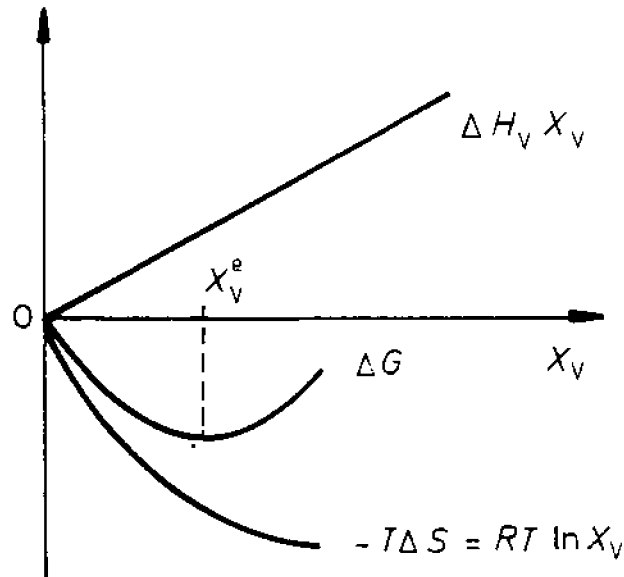


Fig. 1.37 Equilibrium vacancy concentration.

Differentiating Equation 1.55 and making the approximation $X_v \ll 1$ gives

$$\Delta H_v - T\Delta S_v + RT \ln X_v^e = 0$$

Therefore the expression for X_v^e is

$$X_v^e = \exp \frac{\Delta S_v}{R} \cdot \exp \frac{-\Delta H_v}{RT} \quad (1.56)$$

or, putting $\Delta G_v = \Delta H_v - T\Delta S_v$ gives

$$X_v^e = \exp \frac{-\Delta G_v}{RT} \quad (1.57)$$

The first term on the right-hand side of Equation 1.56 is a constant ~ 3 , independent of T , whereas the second term increases rapidly with increasing T . In practice ΔH_v is of the order of 1 eV per atom and X_v^e reaches a value of about 10^{-4} – 10^{-3} at the melting point of the solid.

1.6. The Influence of Interfaces on Equilibrium

The free energy curves that have been drawn so far have been based on the molar free energies of infinitely large amounts of material of a perfect single crystal. Surfaces, grain boundaries and interphase interfaces have been ignored. In real situations these and other crystal defects such as dislocations do exist and raise the free energies of the phases. Therefore the minimum free

energy of an alloy, i.e. the equilibrium state, is not reached until virtually all interfaces and dislocations have been annealed out. In practice such a state is unattainable within reasonable periods of time.

Interphase interfaces can become extremely important in the early stages of phase transformations when one phase, β , say, can be present as very fine particles in the other phase, α , as shown in Fig. 1.38a. If the α phase is acted on by a pressure of 1 atm the β phase is subjected to an extra pressure ΔP due to the curvature of the α/β interface, just as a soap bubble exerts an extra pressure ΔP on its contents. If γ is the α/β interfacial energy and the particles are spherical with a radius r , ΔP is given approximately by

$$\Delta P = \frac{2\gamma}{r}$$

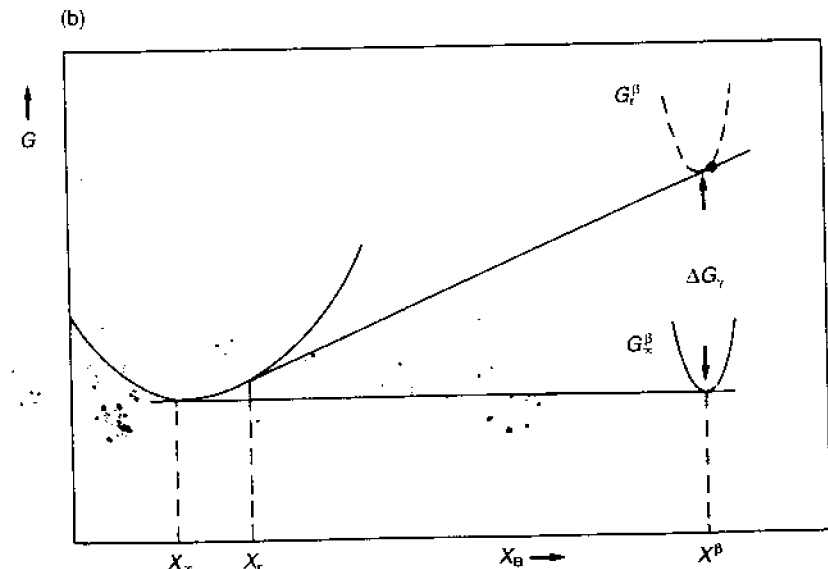
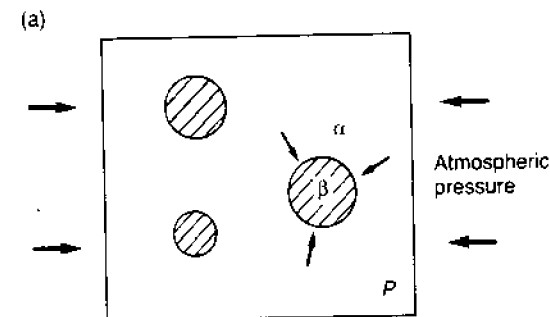


Fig. 1.38 The effect of interfacial energy on the solubility of small particles.

By definition, the Gibbs free energy contains a 'PV' term and an increase of pressure P therefore causes an increase in free energy G . From Equation 1.9 at constant temperature

$$\Delta G = \Delta P \cdot V$$

Therefore the β curve on the molar free energy-composition diagram in Fig. 1.38b will be raised by an amount

$$\Delta G_\gamma = \frac{2\gamma V_m}{r} \quad (1.58)$$

where V_m is the molar volume of the β phase. This free energy increase due to interfacial energy is known as a capillarity effect or the Gibbs-Thomson effect.

The concept of a pressure difference is very useful for spherical liquid particles, but it is less convenient in solids. This is because, as will be discussed in Chapter 3, finely dispersed solid phases are often non-spherical. For illustration, therefore, consider an alternative derivation of Equation 1.58 which can be more easily modified to deal with non-spherical cases³.

Consider a system containing two β particles one with a spherical interface of radius r and the other with a planar interface ($r = \infty$) embedded in an α matrix as shown in Fig. 1.39. If the molar free energy difference between the two particles is ΔG_γ , the transfer of a small quantity (dn mol) of β from the large to the small particle will increase the free energy of the system by a small amount (dG) given by

$$dG = \Delta G_\gamma dn$$

If the surface area of the large particle remains unchanged the increase in free energy will be due to the increase in the interfacial area of the spherical

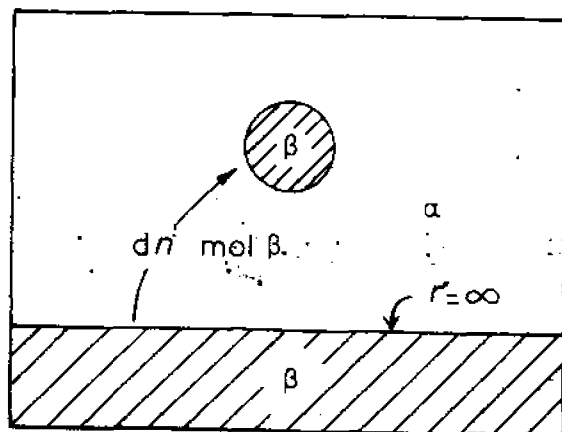


Fig. 1.39 Transfer of dn mol of β from large to a small particle.

particle (dA). Therefore assuming γ is constant

$$dG = \gamma dA$$

Equating these two expressions gives

$$\Delta G_\gamma = \gamma \frac{dA}{dn} \quad (1.59)$$

Since $n = 4\pi r^3/3V_m$ and $A = 4\pi r^2$ it can easily be shown that

$$\frac{dA}{dn} = \frac{dA/dr}{dn/dr} = \frac{2V_m}{r}$$

from which Equation 1.58 can be obtained.

An important practical consequence of the Gibbs-Thomson effect is that the solubility of β in α is sensitive to the size of the β particles. From the common tangent construction in Fig. 1.38b it can be seen that the concentration of solute B in α in equilibrium with β across a curved interface (X_r) is greater than X_∞ , the equilibrium concentration for a planar interface. Assuming for simplicity that the α phase is a regular solution and that the β phase is almost pure B, i.e. $X_B^\beta \sim 1$, Equation 1.52 gives

$$X_\infty = \exp \left\{ -\frac{\Delta G_B + \Omega}{RT} \right\}$$

Similarly X_r can be obtained by using $(\Delta G_B - 2\gamma V_m/r)$ in place of ΔG_B

$$X_r = \exp \left\{ -\frac{\Delta G_B + \Omega - 2\gamma V_m/r}{RT} \right\}$$

Therefore

$$X_r = X_\infty \exp \frac{2\gamma V_m}{RT r} \quad (1.60)$$

and for small values of the exponent

$$X_r \approx X_\infty \left(1 + \frac{2\gamma V_m}{RT r} \right) \quad (1.61)$$

Taking the following typical values: $\gamma = 200 \text{ mJ m}^{-2}$, $V_m = 10^{-5} \text{ m}^3$, $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$, $T = 500 \text{ K}$ gives

$$\frac{X_r}{X_\infty} = 1 + \frac{1}{r(\text{nm})}$$

e.g. for $r = 10 \text{ nm}$ $X_r/X_\infty \sim 1.1$. It can be seen therefore that quite large solubility differences can arise for particles in the range $r = 1\text{--}100 \text{ nm}$. However, for particles visible in the light microscope ($r > 1 \mu\text{m}$) capillarity effects are very small.

1.7 Ternary Equilibrium

Since most commercial alloys are based on at least three components, an understanding of ternary phase diagrams is of great practical importance. The ideas that have been developed for binary systems can be extended to systems with three or more components⁴.

The composition of a ternary alloy can be indicated on an equilateral triangle (the Gibbs triangle) whose corners represent 100% A, B or C as shown in Fig. 1.40. The triangle is usually divided by equidistant lines parallel to the sides marking 10% intervals in atomic or weight per cent. All points on lines parallel to BC contain the same percentage of A, the lines parallel to AC represent constant B concentration, and lines parallel to AB constant C concentrations. Alloys on PQ for example contain 60% A, on RS 30% B, and TU 10% C. Clearly the total percentage must sum to 100%, or expressed as mole fractions

$$X_A + X_B + X_C = 1 \quad (1.62)$$

The Gibbs free energy of any phase can now be represented by a vertical distance from the point in the Gibbs triangle. If this is done for all possible compositions the points trace out the free energy surfaces for all the possible phases, as shown in Fig. 1.41a. The chemical potentials of A, B and C in any phase are then given by the points where the *tangential plane* to the free energy surfaces intersects the A, B and C axes. Figure 1.41a is drawn for a

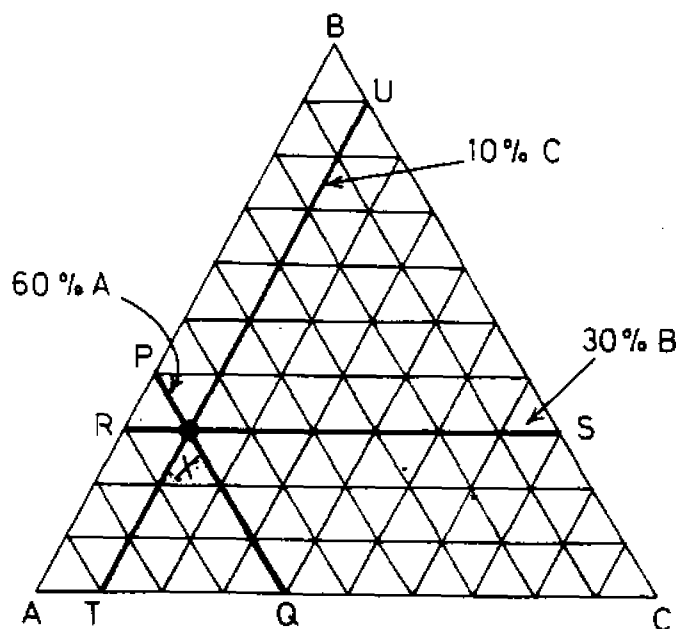


Fig. 1.40 The Gibbs triangle.

system in which the three binary systems AB, BC and CA are simple eutectics. Free energy surfaces exist for three solid phases α , β and γ and the liquid phase, L. At this temperature the liquid phase is most stable for all alloy compositions. At lower temperatures the G^L surface moves upwards and eventually intersects the G^α surface as shown in Fig. 1.41b. Alloys with compositions in the vicinity of the intersection of the two curves consist of α - L at equilibrium. In order for the chemical potentials to be equal in both

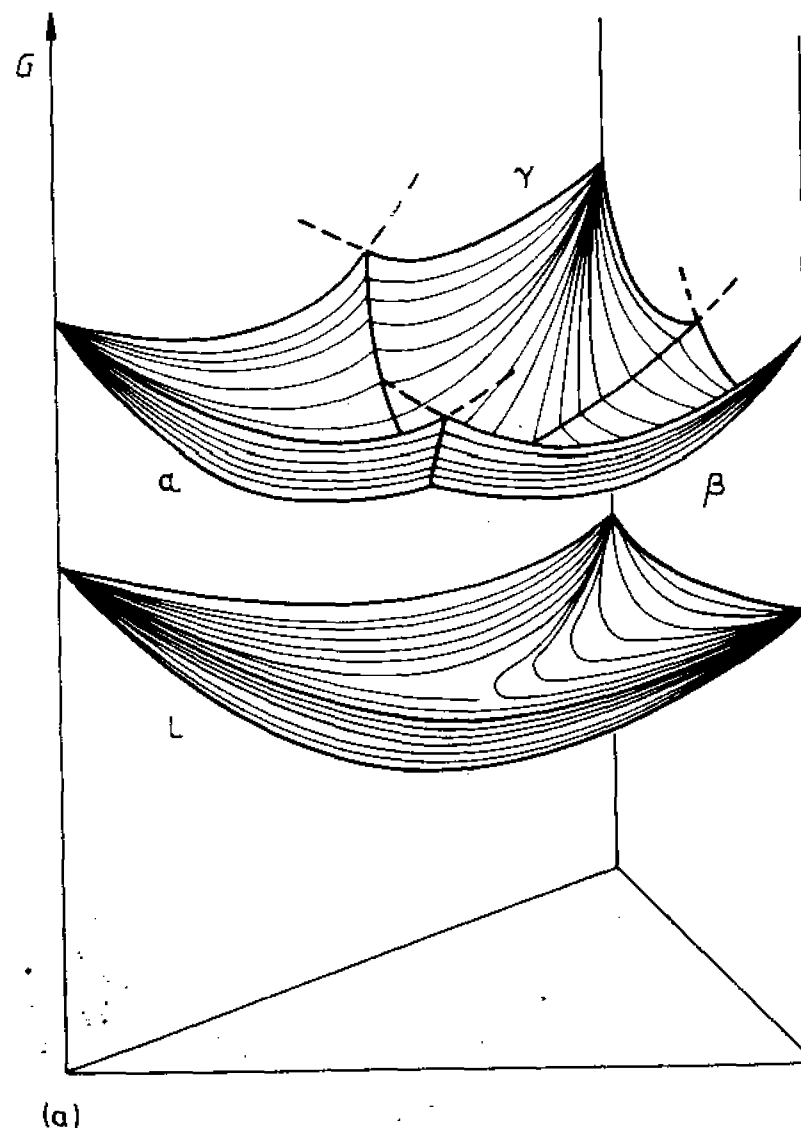


Fig. 1.41 (a) Free energies of a liquid and three solid phases of a ternary system.

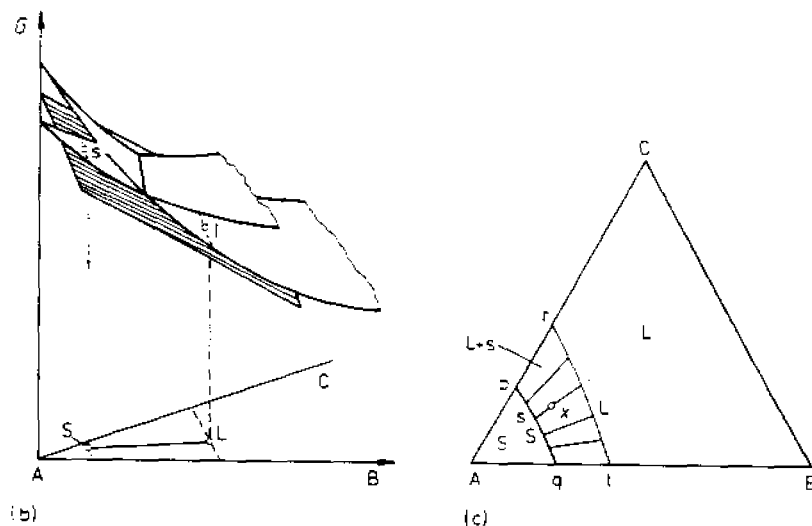


Fig. 1.41 (Cont.) (b) A tangential plane construction to the free energy surfaces defines equilibrium between s and l in the ternary system. (c) Isothermal section through a ternary phase diagram obtained in this way with a two-phase region (L+S) and various tie-lines. The amounts of l and s at point x are determined by the lever rule. (After P. Haasen, *Physical Metallurgy*, Cambridge University Press, Cambridge, 1978.)

phases the compositions of the two phases in equilibrium must be given by points connected by a *common tangential plane*, for example s and l in Fig. 1.41b. These points can be marked on an *isothermal section* of the equilibrium phase diagram as shown in Fig. 1.41c. The lines joining the compositions in equilibrium are known as *tie-lines*. By rolling the tangential plane over the two free energy surfaces a whole series of tie-lines will be generated, such as pr and qt , and the region covered by these tie-lines $pqrt$ is a two-phase region on the phase diagram. An alloy with composition x in Fig. 1.41c will therefore minimize its free energy by separating into solid α with composition s and liquid with composition l . The relative amounts of α and L are simply given by the lever rule. Alloys with compositions within Apq will be a homogeneous α phase at this temperature, whereas alloys within $BCrt$ will be liquid.

On further cooling the free energy surface for the liquid will rise through the other free energy surfaces producing the sequence of isothermal sections shown in Fig. 1.42. In Fig. 1.42f, for example, the liquid is stable near the centre of the diagram whereas at the corners the α , β and γ solid phases are stable. In between are several two-phase regions containing bundles of tie-lines. In addition there are three-phase regions known as *tie-triangles*. The $L - \alpha + \beta$ triangle for example arises because the common tangential plane simultaneously touches the G^α , G^β and G^L surfaces. Therefore any alloy with a composition within the $L + \alpha + \beta$ triangle at this temperature will be in equilibrium as a three-phase mixture with compositions given by the corners

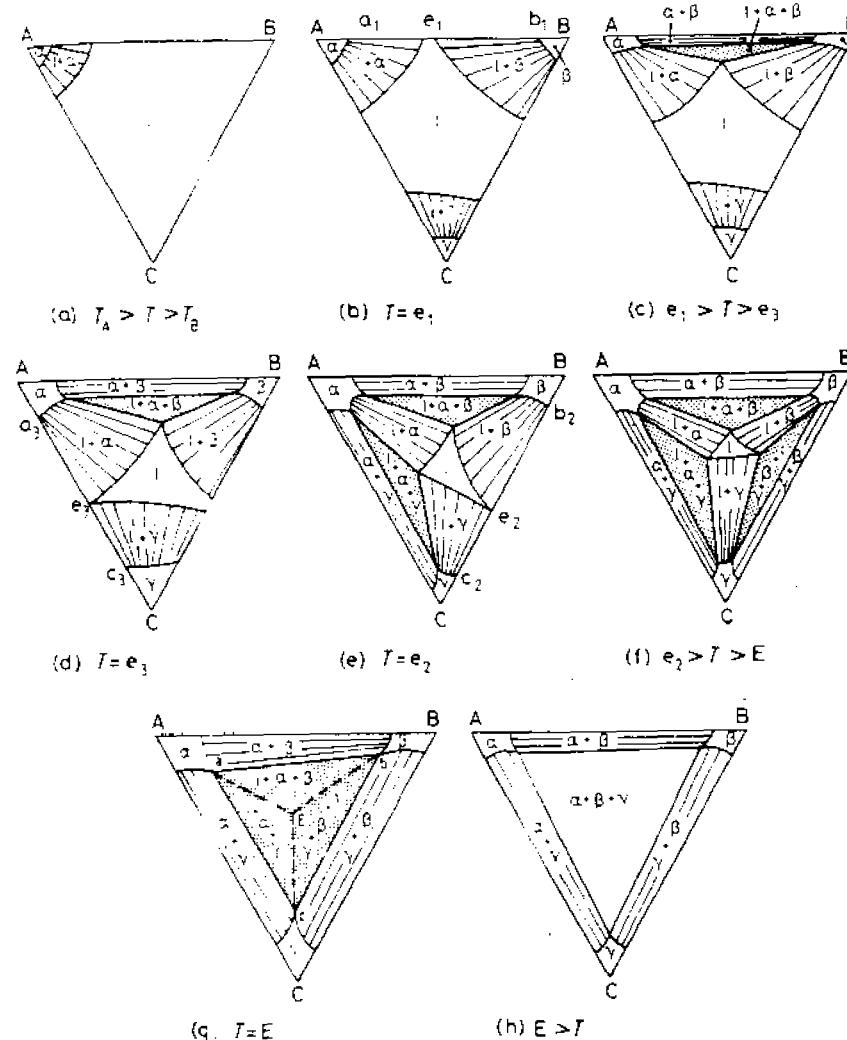


Fig. 1.42 Isothermal sections through Fig 1.41. (After A. Prince, *Alloy Phase Equilibria*, Elsevier, Amsterdam, 1966.)

of the triangle. If the temperature is lowered still further the L region shrinks to a point at which four phases are in equilibrium $L + \alpha + \beta + \gamma$. This is known as the ternary eutectic point and the temperature at which it occurs is the ternary eutectic temperature, Fig. 1.42g. Below this temperature the liquid is no longer stable and an isothermal section contains three two-phase regions and one three-phase tie triangle $\alpha + \beta + \gamma$ as shown in Fig. 1.42h. If isothermal sections are constructed for all temperatures they can be combined into a three-dimensional ternary phase diagram as shown in Fig. 1.44.

In order to follow the course of solidification of a ternary alloy, assuming equilibrium is maintained at all temperatures, it is useful to plot the liquidus

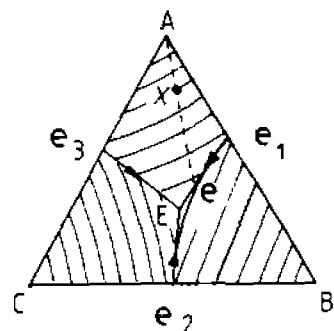


Fig. 1.43 A projection of the liquidus surfaces of Fig. 1.44 onto the Gibbs triangle.

surface contours as shown in Fig. 1.43. During equilibrium freezing of alloy X the liquid composition moves approximately along the line Xe (drawn through A and X) as primary α phase is solidified; then along the eutectic valley eE as both α and β solidify simultaneously. Finally at E , the ternary eutectic point, the liquid transforms simultaneously into $\alpha + \beta + \gamma$. This sequence of events is also illustrated in the perspective drawing in Fig. 1.44.

The phases that form during solidification can also be represented on a vertical section through the ternary phase diagram. Figure 1.45 shows such a section taken through X parallel to AB in Fig. 1.44. It can be seen that on cooling from the liquid phase the alloy first passes into the $L + \alpha$ region, then into $L + \alpha + \beta$, and finally all liquid disappears and the $\alpha + \beta + \gamma$ region is entered, in agreement with the above.

An important limitation of vertical sections is that in general the section will not coincide with the tie-lines in the two-phase regions and so the diagram only shows the phases that exist in equilibrium at different temperatures and *not* their compositions. Therefore they can not be used like binary phase diagrams, despite the superficial resemblance.

1.8 Additional Thermodynamic Relationships for Binary Solutions

It is often of interest to be able to calculate the change in chemical potential ($d\mu$) that results from a change in alloy composition (dX). Considering Fig. 1.46 and comparing triangles it can be seen that

$$\frac{d\mu_A}{dX_B} = \frac{d\mu_B}{dX_A} = \frac{d(\mu_B - \mu_A)}{1} \quad (1.63)$$

and that the slope of the free energy–composition curve is given by

$$\frac{dG}{dX_B} = \frac{\mu_B - \mu_A}{1} \quad (1.64)$$

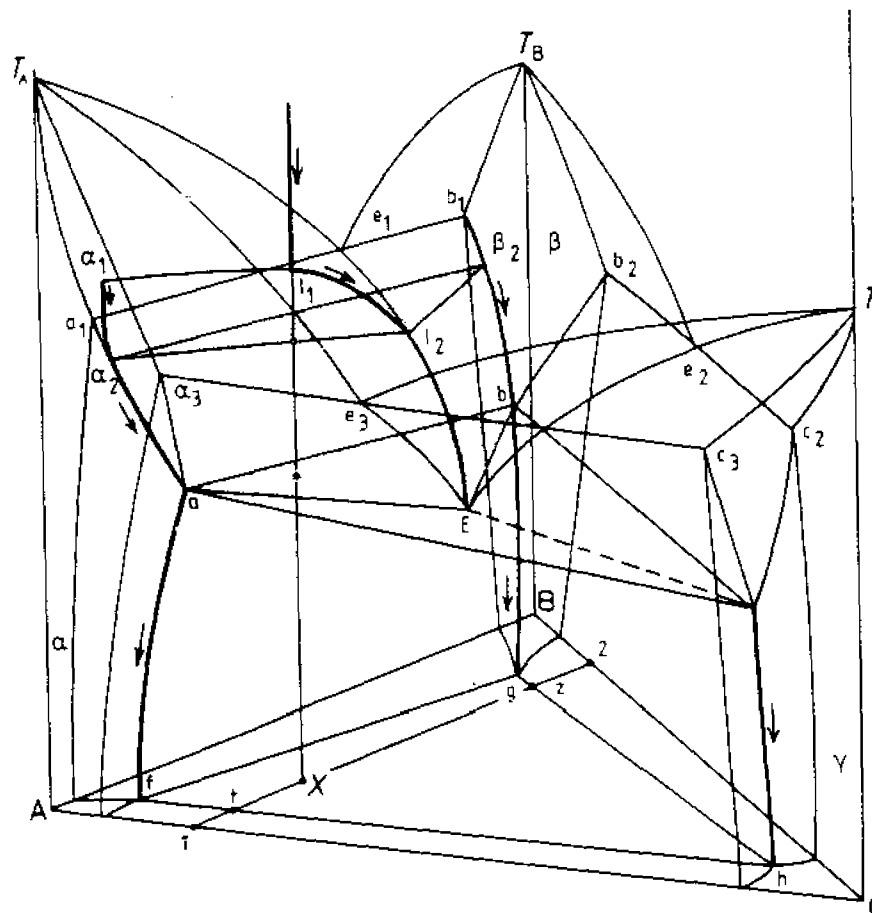


Fig. 1.44 The equilibrium solification of alloy X . (After A. Prince, *Alloy Phase Equilibria*, Elsevier, Amsterdam, 1966.)

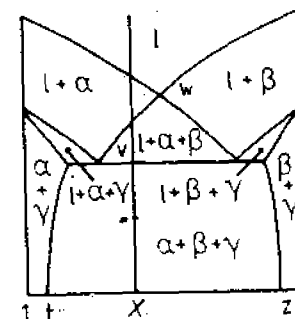


Fig. 1.45 A vertical section between points 1, 2 and X in Fig. 1.44. (After A. Prince, *Alloy Phase Equilibria*, Elsevier, Amsterdam, 1966.)

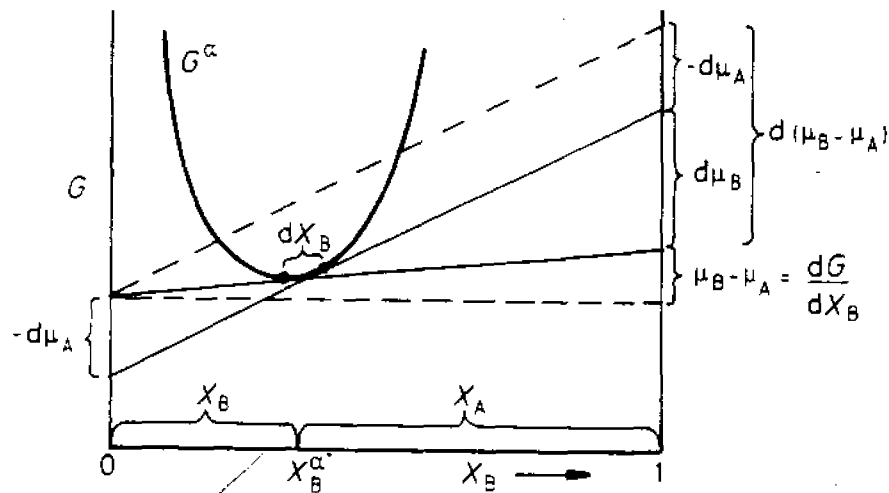


Fig. 1.46 Evaluation of the change in chemical potential due to a change in composition. (After M. Hillert, in *Lectures on the Theory of Phase Transformations*, H.I. Aaronson (Ed.), ©The American Society for Metals and The Metallurgical Society of AIME, New York, 1969.)

Substituting this expression into Equation 1.63 and multiplying throughout by $X_A X_B$ leads to the following equalities:

$$-X_A d\mu_A = X_B d\mu_B = X_A X_B \frac{d^2 G}{dX^2} dX_B \quad (1.65)$$

which are the required equations relating $d\mu_A$, $d\mu_B$ and dX_B . The first equality in this equation is known as the Gibbs-Duhem relationship for a binary solution. Note that the B subscript has been dropped from $d^2 G/dX^2$ as $d^2 G/dX_B^2 \equiv d^2 G/dX_A^2$. For a regular solution differentiation of Equation 1.39 gives

$$\frac{d^2 G}{dX^2} = \frac{RT}{X_A X_B} - 2\Omega \quad (1.56)$$

For an ideal solution $\Omega = 0$ and

$$\frac{d^2 G}{dX^2} = \frac{RT}{X_A X_B} \quad (1.67)$$

Equation 1.65 can be written in a slightly different form by making use of activity coefficients. Combining Equations 1.41 and 1.43 gives

$$\mu_B = G_B + RT \ln \gamma_B X_B \quad (1.68)$$

Therefore

$$\frac{d\mu_B}{dX_B} = \frac{RT}{X_B} \left[1 + \frac{X_B}{\gamma_B} \frac{d\gamma_B}{dX_B} \right] = \frac{RT}{X_B} \left[1 + \frac{d \ln \gamma_B}{d \ln X_B} \right] \quad (1.69)$$

A similar relationship can be derived for $d\mu_A/dX_B$. Equation 1.65 therefore becomes

$$-X_A d\mu_A = X_B d\mu_B = RT \left\{ 1 - \frac{d \ln \gamma_A}{d \ln X_A} \right\} dX_B = RT \left\{ 1 + \frac{d \ln \gamma_B}{d \ln X_B} \right\} dX_B \quad (1.70)$$

Comparing Equations 1.65 and 1.70 gives

$$X_A X_B \frac{d^2 G}{dX^2} = RT \left\{ 1 + \frac{d \ln \gamma_A}{d \ln X_A} \right\} = RT \left\{ 1 - \frac{d \ln \gamma_B}{d \ln X_B} \right\} \quad (1.71)$$

1.9 The Kinetics of Phase Transformations

The thermodynamic functions that have been described in this chapter apply to systems that are in stable or metastable equilibrium. Thermodynamics can therefore be used to calculate the driving force for a transformation, Equation 1.4, but it cannot say how fast a transformation will proceed. The study of how fast processes occur belongs to the science of *kinetics*.

Let us redraw Fig. 1.1 for the free energy of a single atom as it takes part in a phase transformation from an initially metastable state into a state of lower free energy, Fig. 1.47. If G_1 and G_2 are the free energies of the initial and

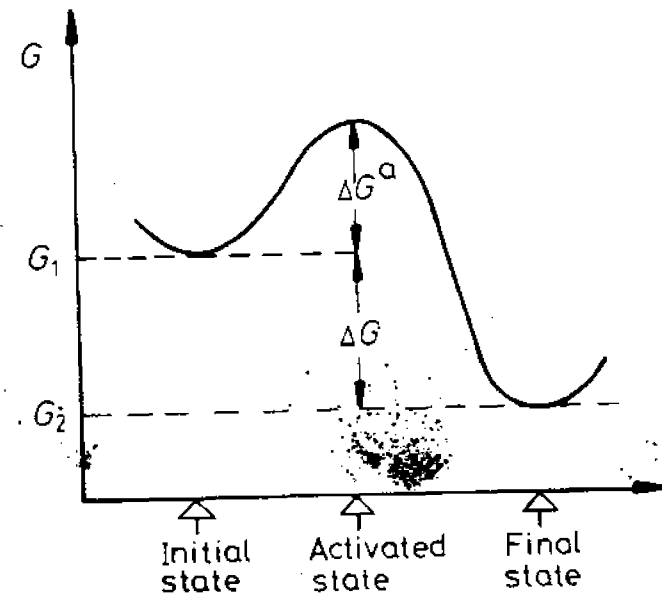


Fig. 1.47 Transformations from initial to final state through an activated state of higher free energy.

final states, the driving force for the transformation will be $\Delta G = G_2 - G_1$. However, before the free energy of the atom can decrease from G_1 to G_2 the atom must pass through a so-called *transition* or *activated* state with a free energy ΔG^a above G_1 . The energies shown in Fig. 1.47 are average energies associated with large numbers of atoms. As a result of the random thermal motion of the atoms the energy of any particular atom will vary with time and occasionally it may be sufficient for the atom to reach the activated state. This process is known as *thermal activation*.

According to kinetic theory, the probability of an atom reaching the activated state is given by $\exp(-\Delta G^a/kT)$ where k is Boltzmann's constant (R/N_A) and ΔG^a is known as the *activation free energy barrier*. The rate at which a transformation occurs will depend on the frequency with which atoms reach the activated state. Therefore we can write

$$\text{rate} \propto \exp\left(-\frac{\Delta G^a}{kT}\right)$$

Putting $\Delta G^a = \Delta H^a - T\Delta S^a$ and changing from atomic to molar quantities enables this equation to be written as

$$\text{rate} \propto \exp\left(-\frac{\Delta H^a}{RT}\right) \quad (1.72)$$

This equation was first derived empirically from the observed temperature dependence of the rate of chemical reactions and is known as the *Arrhenius rate equation*. It is also found to apply to a wide range of processes and transformations in metals and alloys, the simplest of these is the process of diffusion which is discussed in Chapter 2.

References

1. See D.R. Gaskell, *Introduction to Metallurgical Thermodynamics*, p. 342, McGraw-Hill, 1973.
2. For a more detailed treatment of the structures of solid solutions, inter-metallic compounds and intermediate phases see, for example, Chapters 4 and 5 in *Physical Metallurgy*, R.W. Cahn (Ed.), North-Holland, 1974 (second edition).
3. For a treatment of nonspherical precipitates see M. Ferrante and R.D. Doherty, *Acta Metallurgica*, 27: (1979) 1603.
4. See, for example, M. Hillert, *Phase Transformations*, American Society for Metals, Ohio, 1970, Chapter 5.

Further Reading

A.H. Cottrell, *An Introduction to Metallurgy*, Chapter 14 'Alloys' and Chapter 15. 'The phase diagram', Edward Arnold, London, 1967.

- D.R. Gaskell, *Introduction to Metallurgical Thermodynamics*, McGraw-Hill, New York, 1973.
- P. Gordon, *Principles of Phase Diagrams in Materials Systems*, McGraw-Hill, New York, 1968.
- M. Hillert, 'Calculation of Phase Equilibria', Chapter 5 in *Phase Transformations*, American Society for Metals, Ohio, 1970.
- M. Hillert, 'The uses of the Gibbs free energy-composition diagrams', Chapter 1 in *Lectures on the Theory of Phase Transformations*, H.I. Aaronson (Ed.) published by the The Metallurgical Society of AIME, New York, 1975.
- A. Prince, *Alloy Phase Equilibria*, Elsevier, London, 1966.
- G.V. Raynor, 'Phase diagrams and their determination', Chapter 7 in R.W. Cahn (Ed.) *Physical Metallurgy*, North-Holland, 1970.
- F.N. Rhines, *Phase Diagrams in Metallurgy*, McGraw-Hill, New York, 1956.
- P.G. Shewmon, 'Metallurgical thermodynamics', Chapter 6 in R.W. Cahn and P. Haasen (Eds.) *Physical Metallurgy*, North-Holland, 1983.
- A.D. Pelton, 'Phase diagrams', Chapter 7 in R.W. Cahn and P. Haasen (Eds.) *Physical Metallurgy*, North-Holland, 1983.
- R.A. Swalin, *Thermodynamics of Solids*, Wiley, New York, second edition, 1972.
- D.R.F. West, *Ternary Equilibrium Diagrams*, Chapman & Hall, 2nd edition, 1982.
- C.H.P. Lupis, *Chemical Thermodynamics of Materials*, North Holland, 1983.

Exercises

- 1.1 The specific heat of solid copper above 300 K is given by

$$C_p = 22.64 + 6.28 \times 10^{-3} T \quad \text{J mol}^{-1} \text{K}^{-1}$$

By how much does the entropy of copper increase on heating from 300 to 1358 K?

- 1.2 With the aid of Equation 1.11 and Fig. 1.5, draw schematic free energy-pressure curves for pure Fe at 1600, 800, 500 and 300 °C.
- 1.3 Estimate the change in the equilibrium melting point of copper caused by a change of pressure of 10 kbar. The molar volume of copper is $8.0 \times 10^{-6} \text{ m}^3$ for the liquid, and $7.6 \times 10^{-6} \text{ m}^3$ for the solid phase. The latent heat of fusion of copper is $13.05 \text{ kJ mol}^{-1}$. The melting point is 1085 °C.
- 1.4 For a single component system, why do the allotropes stable at high temperatures have higher enthalpies than allotropes stable at low temperatures, e.g. $H(\gamma\text{-Fe}) > H(\alpha\text{-Fe})$?
- 1.5 Determine, by drawing, the number of distinguishable ways of arranging two black balls and two white balls in a square array. Check your answer with Equation 1.24.

- 1.6 By using Equations 1.30 and 1.31, show that the chemical potentials of A and B can be obtained by extrapolating the tangent to the G - X curve to $X_A = 0$ and $X_B = 0$.
- 1.7 Derive Equation 1.40 from 1.31 and 1.39.
- 1.8 15 g of gold and 25 g of silver are mixed to form a single-phase ideal solid solution.
- How many moles of solution are there?
 - What are the mole fractions of gold and silver?
 - What is the molar entropy of mixing?
 - What is the *total* entropy of mixing?
 - What is the molar free energy change at 500 °C?
 - What are the chemical potentials of Au and Ag at 500 °C taking the free energies of pure Au and Ag as zero?
 - By how much will the free energy of the solution change at 500 °C if one Au atom is added? Express your answer in eV/atom.
- 1.9 In the Fe-C system Fe_3C is only a metastable phase, whilst graphite is the most stable carbon-rich phase. By drawing schematic free energy-composition diagrams show how the Fe-graphite phase diagram compares to the Fe- Fe_3C phase diagram from 0 to 2 wt% Fe. Check your answer with the published phase diagram in the *Metals Handbook* for example.
- 1.10 Consider a multicomponent system A, B, C . . . containing several phases α , β , γ . . . at equilibrium. If a small quantity of A ($dn_A \text{ mol}$) is taken from the α phase and added to the β phase at constant T and P what are the changes in the free energies of the α and β phases, dG^α and dG^β ? Since the overall mass and composition of the system is unchanged by the above process the total free energy change $dG = dG^\alpha + dG^\beta = 0$. Show, therefore, that $\mu_A^\alpha = \mu_A^\beta$. Repeating for other pairs of phases and other components gives the general equilibrium conditions, Equation 1.48.
- 1.11 For aluminium $\Delta H_v = 0.8 \text{ eV atom}^{-1}$ and $\Delta S_v/R = 2$. Calculate the equilibrium vacancy concentration at 660 °C (T_m) and 25 °C.
- 1.12 The solid solubility of silicon in aluminium is 1.25 atomic % at 550 °C and 0.46 atomic % at 450 °C. What solubility would you expect at 200 °C? Check your answer by reference to the published phase diagram.
- 1.13 The metals A and B form an ideal liquid solution but are almost immiscible in the solid state. The entropy of fusion of both A and B is $8.4 \text{ J mol}^{-1} \text{ K}^{-1}$ and the melting temperatures are 1500 and 1300 K respectively. Assuming that the specific heats of the solid and liquid are identical calculate the eutectic composition and temperature in the A-B phase diagram.
- 1.14 Write down an equation that shows by how much the molar free energy of solid Cu is increased when it is present as a small sphere of radius r in liquid Cu. By how much must liquid Cu be cooled below T_m before a

solid particle of Cu can grow if the particle *diameter* is (i) 2 μm , (ii) 2 nm (20 Å)? (Cu: $T_m = 1085 \text{ °C} = 1358 \text{ K}$. Atomic weight 63.5. Density 8900 kg m^{-3} . Solid/liquid interfacial energy $\gamma = 0.144 \text{ J m}^{-2}$. Latent heat of melting $L = 13\,300 \text{ J mol}^{-1}$.)

- 1.15 Suppose a ternary alloy containing 40 atomic % A, 20 atomic % B, 40 atomic % C solidifies through a ternary eutectic reaction to a mixture of α , β and γ with the following compositions: 80 atomic % A, 5 atomic % B, 15 atomic % C; 70 atomic % B, 10 atomic % A, 20 atomic % C; and 20 atomic % B, 10 atomic % A, 70 atomic % C. What will be the mole fractions of α , β and γ in the microstructure?
- 1.16 Show that a general expression for the chemical potential of a component in solution is given by

$$\mu_A = G_A^0 + S_A(T_0 - T) - RT \ln \gamma_A X_A + (P - P_0)V_m$$

where G_A^0 is the free energy of pure A at temperature T_0 and pressure P_0 , S_A is the entropy of A, R is the gas constant, γ_A the activity coefficient for A, X_A the mole fraction in solution, V_m is the molar volume which is assumed to be constant. Under what conditions is the above equation valid?

2

Diffusion

The previous chapter was mainly concerned with stable or equilibrium arrangements of atoms in an alloy. The study of phase transformations concerns those mechanisms by which a system attempts to reach this state and how long it takes. One of the most fundamental processes that controls the rate at which many transformations occur is the diffusion of atoms.

The reason why diffusion occurs is always so as to produce a decrease in Gibbs free energy. As a simple illustration of this consider Fig. 2.1. Two blocks of the same A–B solid solution, but with different compositions, are welded together and held at a temperature high enough for long-range diffusion to occur. If the molar free energy diagram of the alloy is as shown in Fig. 2.1b, the molar free energy of each part of the alloy will be given by G_1 and G_2 , and initially the total free energy of the welded block will be G_3 . However, if diffusion occurs as indicated in Fig. 2.1a so as to eliminate the concentration differences, the free energy will decrease towards G_4 , the free energy of a homogeneous alloy. Thus, in this case, a decrease in free energy is produced by A and B atoms diffusing away from the regions of high concentration to that of low concentration, i.e. down the concentration gradients. However, this need not always be the case as was indicated in Section 1.4. In alloy systems that contain a miscibility gap the free energy curves can have a negative curvature at low temperatures. If the free energy curve and composition for the A–B alloy shown in Fig. 2.1a were as drawn in Fig. 2.1d the A and B atoms would diffuse towards the regions of high concentration, i.e. up the concentration gradients, as shown in Fig. 2.1c. However, this is still the most natural process as it reduces the free energy from G_3 towards G_4 again.

As can be seen in Fig. 2.1e and f the A and B atoms are diffusing from regions where the chemical potential is high to regions where it is low, i.e. down the chemical potential gradient in both cases. In practice the first case mentioned above is far more common than the second case, and it is usually assumed that diffusion occurs down concentration gradients. However, it can be seen that this is only true under special circumstances and for this reason it is strictly speaking better to express the driving force for diffusion in terms of a chemical potential gradient. Diffusion ceases when the chemical potentials of all atoms are everywhere the same and the system is in equilibrium. However, since case 1 above is mainly encountered in practice and because concentration differences are much easier to measure than chemical potential differences, it is nevertheless more convenient to relate diffusion to concentration gradients. The remainder of this chapter will thus be mainly concerned with this approach to diffusion.

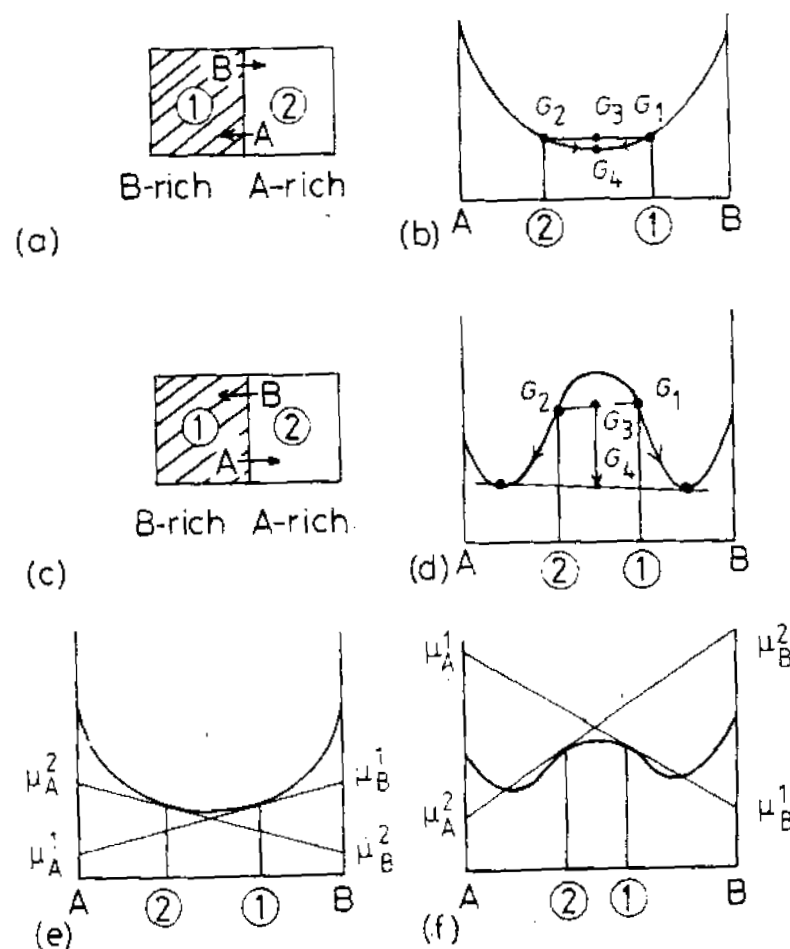


Fig. 2.1 Free energy and chemical potential changes during diffusion. (a) and (b) 'down-hill' diffusion. (c) and (d) 'up-hill' diffusion. (e) $\mu_A^2 > \mu_A^1$ therefore A atoms move from (2) to (1). $\mu_B^1 > \mu_B^2$ therefore B atoms move from (1) to (2). (f) $\mu_A^1 > \mu_A^2$ therefore A atoms move from (1) to (2). $\mu_B^2 > \mu_B^1$ therefore B atoms move from (2) to (1).

2.1 Atomic Mechanisms of Diffusion

There are two common mechanisms by which atoms can diffuse through a solid and the operative mechanism depends on the type of site occupied in the lattice. *Substitutional* atoms usually diffuse by a *vacancy* mechanism whereas the smaller *interstitial* atoms migrate by forcing their way between the larger atoms, i.e. *interstitially*.

Normally a substitutional atom in a crystal oscillates about a given site and is surrounded by neighbouring atoms on similar sites. The mean vibrational

energy possessed by each atom is given by $3 kT$, and therefore increases in proportion to the absolute temperature. Since the mean frequency of vibration is approximately constant the vibrational energy is increased by increasing the amplitude of the oscillations. Normally the movement of a substitutional atom is limited by its neighbours and the atom cannot move to another site. However, if an adjacent site is vacant it can happen that a particularly violent oscillation results in the atom jumping over on to the vacancy. This is illustrated in Fig. 2.2. Note that in order for the jump to occur the shaded atoms in Fig. 2.2b must move apart to create enough space for the migrating atom to pass between. Therefore the probability that any atom will be able to jump into a vacant site depends on the probability that it can acquire sufficient vibrational energy. The rate at which any given atom is able to migrate through the solid will clearly be determined by the frequency with which it encounters a vacancy and this in turn depends on the concentration of vacancies in the solid. It will be shown that both the probability of jumping and the concentration of vacancies are extremely sensitive to temperature.

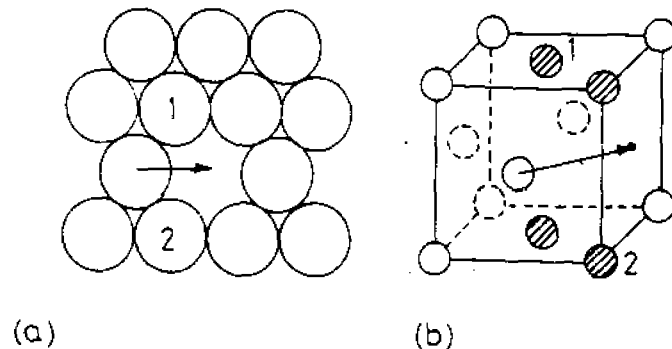


Fig. 2.2 Movement of an atom into an adjacent vacancy in an fcc lattice. (a) A close-packed plane. (b) A unit cell showing the four atoms (shaded) which must move before the jump can occur. (After P.G. Shewmon, *Diffusion in Solids*, McGraw-Hill, New York, 1963.)

When a solute atom is appreciably smaller in diameter than the solvent, it occupies one of the interstitial sites between the solvent atoms. In fcc materials the interstitial sites are midway along the cube edges or, equivalently, in the middle of the unit cell, Fig. 2.3a. These are known as octahedral sites since the six atoms around the site form an octahedron. In the bcc lattice the interstitial atoms also often occupy the octahedral sites which are now located at edge-centring or face-centring positions as shown in Fig. 2.3b.

Usually the concentration of interstitial atoms is so low that only a small fraction of the available sites is occupied. This means that each interstitial atom is always surrounded by vacant sites and can jump to another position as often as its thermal energy permits it to overcome the strain energy barrier to migration, Fig. 2.4.

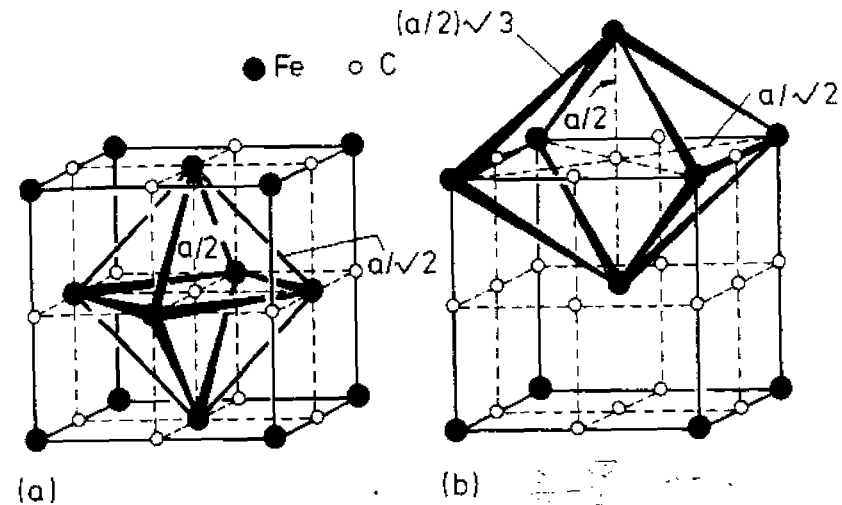


Fig. 2.3 (a) Octahedral interstices (0) in an fcc crystal. (b) Octahedral interstices in a bcc crystal. (After P. Haasen, *Physical Metallurgy*, Cambridge University Press, Cambridge, 1978.)

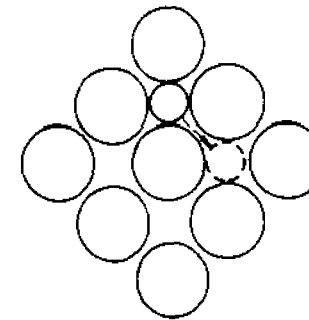


Fig. 2.4 A $\{100\}$ plane in an fcc lattice showing the path of an interstitial atom diffusing by the interstitial mechanism.

2.2 Interstitial Diffusion

2.2.1 Interstitial Diffusion as a Random Jump Process

Let us consider first a simple model of a dilute interstitial solid solution where the parent atoms are arranged on a simple cubic lattice and the solute B atoms fit perfectly into the interstices without causing any distortion of the parent lattice. We assume that the solution is so dilute that every interstitial atom is surrounded by six vacant interstitial sites. If the concentration of B varies in one dimension (x) through the solution (see Fig. 2.5) the B atoms can diffuse throughout the material until their concentration is the same everywhere. The problem to be considered then, concerns how this diffusion is related to the random jump characteristics of the interstitial atoms.

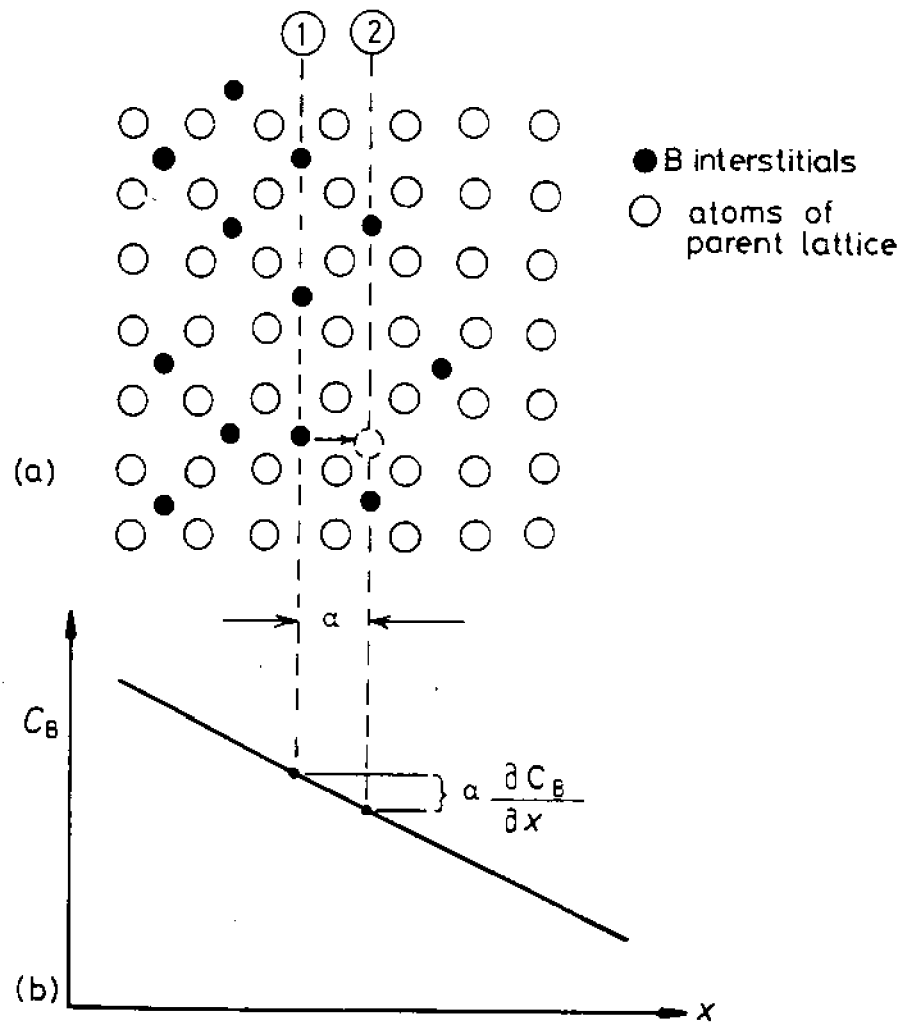


Fig. 2.5 Interstitial diffusion by random jumps in a concentration gradient.

To answer this question consider the exchange of atoms between two adjacent atomic planes such as (1) and (2) in Fig. 2.5a. Assume that on average an interstitial atom jumps Γ_B times per second (Γ = Greek capital gamma) and that each jump is in a random direction, i.e. there is an equal probability of the atom jumping to every one of the six adjacent sites. If plane (1) contains n_1 B-atoms per m^2 the number of atoms that will jump from plane (1) to (2) in 1 s (J) will be given by:

$$J_B = \frac{1}{6} \Gamma_B n_1 \quad \text{atoms m}^{-2} \text{ s}^{-1} \quad (2.1)$$

During the same time the number of atoms that jump from plane (2) to (1),

assuming Γ_B is independent of concentration, is given by:

$$\bar{J}_B = \frac{1}{6} \Gamma_B n_2 \quad \text{atoms m}^{-2} \text{ s}^{-1}$$

Since $n_1 > n_2$ there will be a net flux of atoms from left to right given by:

$$J_B = \bar{J}_B - \bar{J}_B = \frac{1}{6} \Gamma_B (n_1 - n_2) \quad (2.2)$$

where n_1 and n_2 are related to the concentration of B in the lattice. If the separation of planes (1) and (2) is α the concentration of B at the position of plane (1) $C_B(1) = n_1/\alpha$ atoms m^{-3} . Likewise $C_B(2) = n_2/\alpha$. Therefore $(n_1 - n_2) = \alpha(C_B(1) - C_B(2))$ and from Fig. 2.5b it can be seen that $C_B(1) - C_B(2) = -\alpha(\partial C_B/\partial x)$. Substituting these equations into Equation 2.2 gives:

$$J_B = -\left(\frac{1}{6} \Gamma_B \alpha^2\right) \frac{\partial C_B}{\partial x} \quad \text{atoms m}^{-2} \text{ s}^{-1}$$

The partial derivative $\partial C_B/\partial x$ has been used to indicate that the concentration gradient can change with time. Thus in the presence of a concentration gradient the random jumping of individual atoms produces a *net* flow of atoms down the concentration gradient.

Substituting

$$D_B = \frac{1}{6} \Gamma_B \alpha^2 \quad (2.3)$$

yields:

$$J_B = -D_B \frac{\partial C_B}{\partial x} \quad (2.4)$$

This equation is identical to that proposed by Fick in 1855 and is usually known as *Fick's first law of diffusion*. D_B is known as the *intrinsic diffusivity* or the *diffusion coefficient* of B, and has units [$\text{m}^2 \text{s}^{-1}$]. The units for J are [quantity $\text{m}^{-2} \text{s}^{-1}$] and for $\partial C/\partial x$ [quantity m^{-4}], where the unit of quantity can be in terms of atoms, moles, kg, etc. as long it is the same for J and C .

When the jumping of B atoms is truly random with a frequency independent of concentration, D_B is given by Equation 2.3 and is also a constant independent of concentration. Although this equation for D_B was derived for interstitial diffusion in a simple cubic lattice it is equally applicable to any *randomly* diffusing atom in any *cubic* lattice provided the correct substitution for the jump distance α is made. In non-cubic lattices the probability of jumps in different crystallographic directions is not equal and D varies with direction. Atoms in hexagonal lattices, for example, diffuse at different rates parallel and perpendicular to the basal plane.

The condition that the atomic jumps occur completely randomly and independently of concentration is usually not fulfilled in real alloys. Nevertheless it is found from experiment that Fick's first law is still applicable, though only if the diffusion coefficient D is made to vary with composition. For example the diffusion coefficient for carbon in fcc-Fe at 1000 °C is $2.5 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ at 0.15 wt% C, but it rises to $7.7 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ in solutions containing 1.4 wt% C. The reason for the increase of D_C^γ with concentration is that the C atoms strain the Fe lattice thereby making diffusion easier as the amount of strain increases.

As an example of the use of Equation 2.3 the following data can be used to estimate the jump frequency of a carbon atom in γ -Fe at 1000 °C. The lattice parameter of γ -Fe is $\sim 0.37 \text{ nm}$ thus the jump distance $\alpha = 0.37/\sqrt{2} = 0.26 \text{ nm}$ (2.6 Å). Assuming $D = 2.5 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$, leads to the result that $\Gamma = 2 \times 10^9 \text{ jumps s}^{-1}$. If the vibration frequency of the carbon atoms is $\sim 10^{13}$, then only about one attempt in 10^4 results in a jump from one site to another.

It is also interesting to consider the diffusion process from the point of view of a single diffusing atom. If the direction of each new jump is independent of the direction of the previous jump the process is known as a random walk. For a random walk in three dimensions it can be shown¹ that after n steps of length α the 'average' atom will be displaced by a *net* distance $\alpha\sqrt{n}$ from its original position. (This is more precisely the root mean square displacement after n steps.) Therefore after a time t the average atom will have advanced a radial distance r from the origin, where

$$r = \alpha\sqrt{\Gamma t} \quad (2.5)$$

Substituting Equation 2.3 for Γ gives

$$r = 2.4\sqrt{(Dt)} \quad (2.6)$$

It will be seen that the distance $\sqrt{(Dt)}$ is a very important quantity in diffusion problems.

For the example of carbon diffusing in γ -Fe above, in 1 s each carbon atom will move a total distance of $\sim 0.5 \text{ m}$ but will only reach a *net* displacement of $\sim 10 \mu\text{m}$. It is obvious that very few of the atom jumps provide a useful contribution to the total diffusion distance.

2.2.2 Effect of Temperature—Thermal Activation

Let us now take a closer look at the actual jump process for an interstitial atom as in Fig. 2.6a. Due to the thermal energy of the solid all the atoms will be vibrating about their rest positions and occasionally a particularly violent oscillation of an interstitial atom, or some chance coincidence of the movements of the matrix and interstitial atoms, will result in a jump. Since the diffusion coefficient is closely related to the frequency of such jumps, Γ , it is of

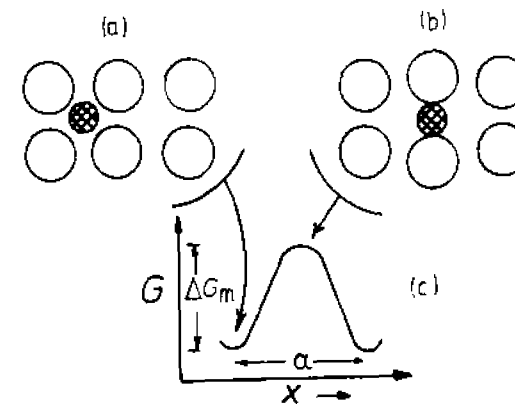


Fig. 2.6 Interstitial atom. (a) in equilibrium position, (b) at the position of maximum lattice distortion. (c) Variation of the free energy of the lattice as a function of the position of interstitial. (After P.G. Shewmon, in *Physical Metallurgy*, 2nd edn., R.W. Cahn (Ed.), North-Holland, Amsterdam, 1974.)

interest to know the factors controlling Γ and the effect of raising the temperature of the system.

The rest positions of the interstitial atoms are positions of minimum potential energy. In order to move an interstitial atom to an adjacent interstice the atoms of the parent lattice must be forced apart into higher energy positions as shown in Fig. 2.6b. The work that must be done to accomplish this process causes an increase in the free energy of the system by ΔG_m (m refers to migration) as shown in Fig. 2.6c. ΔG_m is known as the *activation energy* for the migration of the interstitial atom. In any system in thermal equilibrium the atoms are constantly colliding with one another and changing their vibrational energy. On average, the fraction of atoms with an energy of ΔG or more than the mean energy is given by $\exp(-\Delta G/RT)$. Thus if the interstitial atom in Fig. 2.6a is vibrating with a mean frequency ν in the x direction it makes ν attempts per second to jump into the next site and the fraction of these attempts that are successful is given by $\exp(-\Delta G_m/RT)$. Now the atom is randomly vibrating in three-dimensional space, and if it is surrounded by z sites to which it can jump the jump frequency is given by

$$\Gamma_B = z\nu \exp \frac{-\Delta G_m}{RT} \quad (2.7)$$

ΔG_m can be considered to be the sum of a large activation enthalpy ΔH_m and a small activation entropy term $-T\Delta S_m$.

Combining this expression with Equation 2.3 gives the diffusion coefficient as

$$D_B = \left[\frac{1}{6} \alpha^2 z \nu \exp \frac{\Delta S_m}{R} \right] \exp \frac{-\Delta H_m}{RT} \quad (2.8)$$

This can be simplified to an Arrhenius-type equation, that is

$$D_B = D_{B0} \exp \frac{-Q_{ID}}{RT} \quad (2.9)$$

where

$$D_{B0} = \frac{1}{6} \alpha^2 \nu \exp \frac{\Delta S_m}{R} \quad (2.10)$$

and

$$Q_{ID} = \Delta H_m \quad (2.11)$$

The terms that are virtually independent of temperature have been grouped into a single material constant D_0 . Therefore D or Γ increases exponentially with temperature at a rate determined by the activation enthalpy Q_{ID} (ID refers to Interstitial Diffusion). Equation 2.9 is found to agree with experimental measurements of diffusion coefficients in substitutional as well as interstitial diffusion. In the case of interstitial diffusion it has been shown that the activation enthalpy Q is only dependent on the activation energy barrier to the movement of interstitial atoms from one site to another.

Some experimental data for the diffusion of various interstitials in bcc-Fe are given in Table 2.1. Note that the activation enthalpy for interstitial diffusion increases as the size of the interstitial atom increases. (The atomic diameters decrease in the order C, N, H.) This is to be expected since smaller atoms cause less distortion of the lattice during migration.

A convenient graphical representation of D as a function of temperature can be obtained writing Equation 2.9 in the form

$$\log D = \log D_0 - \frac{Q}{2.3 R} \left(\frac{1}{T} \right) \quad (2.12)$$

Thus if $\log D$ is plotted against $(1/T)$ a straight line is obtained with a slope equal to $-(Q/2.3 R)$ and an intercept on the $\log D$ axis at $\log D_0$, see Fig. 2.7.

Table 2.1 Experimental Diffusion Data for Interstitials in Ferritic (bcc) Iron

Solute	$D_0/\text{mm}^2 \text{ s}^{-1}$	$Q/\text{kJ mol}^{-1}$	Ref.
C	2.0	84.1	2
N	0.3	76.1	3
H	0.1	13.4	4

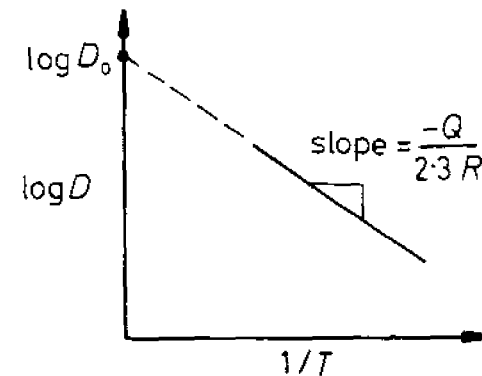


Fig. 2.7 The slope of $\log D$ v. $1/T$ gives the activation energy for diffusion Q .

2.2.3 Steady-State Diffusion

The simplest type of diffusion to deal with is when a steady state exists, that is when the concentration at every point does not change with time. For example consider a thin-walled pressure vessel containing hydrogen. The concentration of hydrogen at the inner surface of the vessel will be maintained at a level C_H depending on the pressure in the vessel, while the concentration at the outer surface is reduced to zero by the escape of hydrogen to the surroundings. A steady state will eventually be reached when the concentration everywhere reaches a constant value. Provided D_H is independent of concentration there will be a single concentration gradient in the wall given by

$$\frac{\partial C}{\partial x} = \frac{0 - C_H}{l}$$

where l is the wall thickness. On this basis the flux through the wall is given by

$$J_H = \frac{D_H C_H}{l} \quad (2.13)$$

2.2.4 Nonsteady-State Diffusion

In most practical situations steady-state conditions are not established, i.e. concentration varies with both distance and time, and Fick's first law can no longer be used. For simplicity let us consider the situation shown in Fig. 2.8a where a concentration profile exists along one dimension (x) only. The flux at any point along the x -axis will depend on the local value of D_B and $\partial C_B / \partial x$ as shown in Fig. 2.8b. In order to calculate how the concentration of B at any point varies with time consider a narrow slice of material with an area A and a thickness δx as shown in Fig. 2.8c.

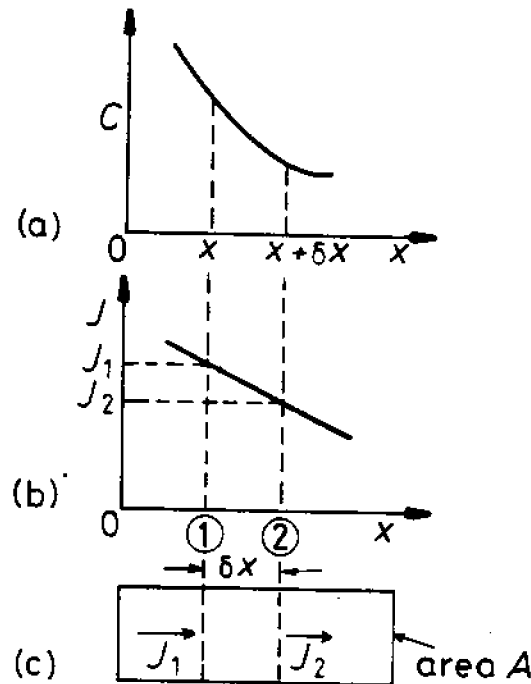


Fig. 2.8 The derivation of Fick's second law.

The number of interstitial B atoms that diffuse into the slice across plane (1) in a small time interval δt will be $J_1 A \delta t$. The number of atoms that leave the thin slice during this time, however, is only $J_2 A \delta t$. Since $J_2 < J_1$, the concentration of B within the slice will have increased by

$$\delta C_B = \frac{(J_1 - J_2) A \delta t}{A \delta x} \quad (2.14)$$

But since δx is small,

$$J_2 = J_1 + \frac{\partial J}{\partial x} \delta x \quad (2.15)$$

and in the limit as $\delta t \rightarrow 0$ these equations give

$$\frac{\partial C_B}{\partial t} = -\frac{\partial J_B}{\partial x} \quad (2.16)$$

Substituting Fick's first law gives

$$\frac{\partial C_B}{\partial t} = \frac{\partial}{\partial x} \left(D_B \frac{\partial C_B}{\partial x} \right) \quad (2.17)$$

which is referred to as *Fick's second law*. If variations of D_B with concentration can be ignored this equation can be simplified to

$$\frac{\partial C_B}{\partial t} = D_B \frac{\partial^2 C_B}{\partial x^2} \quad (2.18)$$

These equations relate the rate of change of composition with time to the concentration profile $C_B(x)$. Equation 2.18 has a simple graphical interpretation as $\partial^2 C_B / \partial x^2$ is the *curvature* of the C_B versus x curve. If the concentration profile appears as shown in Fig. 2.9a it has a positive curvature everywhere and the concentration at all points on such a curve will increase with time ($\partial C_B / \partial t$ positive). When the curvature is negative as in Fig. 2.9b C_B decreases with time ($\partial C_B / \partial t$ negative).

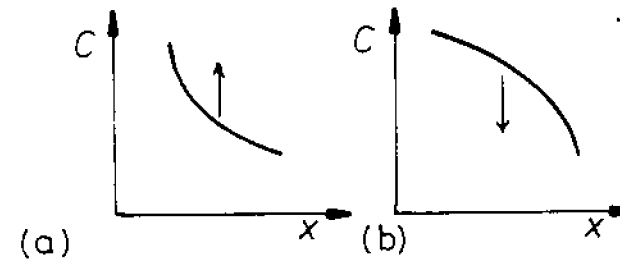


Fig. 2.9 (a) $\partial^2 C / \partial x^2 > 0$ all concentrations increase with time. (b) $\partial^2 C / \partial x^2 < 0$ all concentrations decrease with time.

2.2.5 Solutions to the Diffusion Equation

Two solutions will be considered which are of practical importance. One concerns the situation which is encountered in homogenization heat treatments, and the other is encountered, for example, in the carburization of steel.

Homogenization

It is often of interest to be able to calculate the time taken for an inhomogeneous alloy to reach complete homogeneity, as for example in the elimination of segregation in castings.

The simplest composition variation that can be solved mathematically is if C_B varies sinusoidally with distance in one dimension as shown in Fig. 2.10. In this case B atoms diffuse down the concentration gradients, and regions with negative curvature, such as between $x = 0$ and $x = l$, decrease in concentration, while regions between $x = l$ and $2l$ increase in concentration. The curvature is zero at $x = 0, l, 2l$, so the concentrations at these points remain unchanged with time. Consequently the concentration profile after a certain time reduces to that indicated by the dashed line in Fig. 2.10.

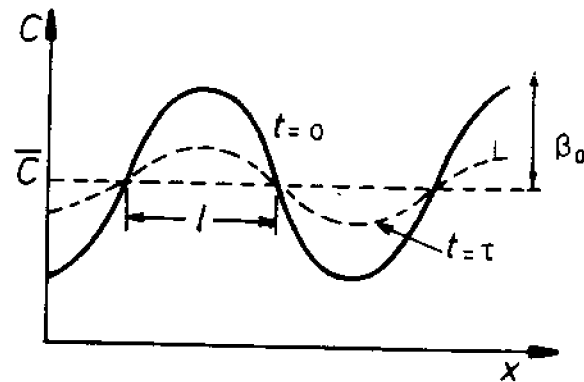


Fig. 2.10 The effect of diffusion on a sinusoidal variation of composition.

At time $t = 0$ the concentration profile is given by

$$C = \bar{C} + \beta_0 \sin \frac{\pi x}{l} \quad (2.19)$$

where $\bar{C}$ is the mean composition, and β_0 is the amplitude of the initial concentration profile. Assuming D_B is independent of concentration the solution of Equation 2.18 that satisfies this initial condition is

$$C = \bar{C} + \beta_0 \sin \left(\frac{\pi x}{l} \right) \exp \frac{-t}{\tau} \quad (2.20)$$

where τ is a constant called the *relaxation time* and is given by:

$$\tau = \frac{l^2}{\pi^2 D_B} \quad (2.21)$$

Thus the amplitude of the concentration profile after a time t (β) is given by C at $x = l/2$, i.e.

$$\beta = \beta_0 \exp \frac{-t}{\tau} \quad (2.22)$$

In other words, the amplitude of the concentration profile decreases exponentially with time and after a sufficiently long time approaches zero so that $C = \bar{C}$ everywhere. The rate at which this occurs is determined by the relaxation time τ . After a time $t = \tau$, $\beta = \beta_0/e$, that is, the amplitude has decreased to $1/2.72$ of its value at $t = 0$. The solute distribution at this stage would therefore appear as shown by the dashed line in Fig. 2.10. After a time $t = 2\tau$ the amplitude is reduced by a total of $1/e^2$, i.e. by about one order of magnitude. From Equation 2.21 it can be seen that the rate of homogenization increases rapidly as the wavelength of the fluctuations decreases.

The initial concentration profile will not usually be sinusoidal, but in general *any* concentration profile can be considered as the sum of an infinite

series of sine waves of varying wavelength and amplitude, and each wave decays at a rate determined by its own τ . Thus the short wavelength terms die away very rapidly and the homogenization will ultimately be determined by τ for the longest wavelength component.

The Carburization of Steel

The aim of carburization is to increase the carbon concentration in the surface layers of a steel product in order to achieve a harder wear-resistant surface. This is usually done by holding the steel in a gas mixture containing CH_4 and/or CO at a temperature where it is austenitic. By controlling the relative proportions of the two gases the concentration of carbon at the surface of the steel in equilibrium with the gas mixture can be maintained at a suitable constant value. At the same time carbon continually diffuses from the surface into the steel.

The concentration profiles that are obtained after different times are shown in Fig. 2.11. An analytical expression for these profiles can be obtained by solving Fick's second law using the boundary conditions: C_B (at $x = 0$) = C_s and $C_B(\infty) = C_0$, the original carbon concentration of the steel. The specimen is considered to be infinitely long. In reality the diffusion coefficient of carbon in austenite increases with increasing concentration, but an approximate solution can be obtained by taking an average value and this gives the simple solution

$$C = C_s - (C_s - C_0) \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (2.23)$$

Where 'erf' stands for error function which is an indefinite integral defined by the equation

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-y^2) dy$$

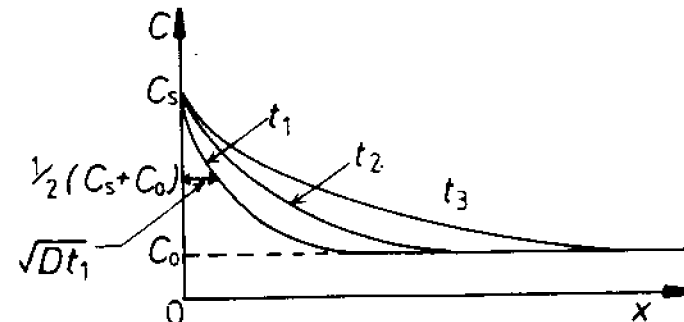


Fig. 2.11 Concentration profiles at successive times ($t_3 > t_2 > t_1$) for diffusion into a semi-infinite bar when the surface concentration C_s is maintained constant.

The function is shown graphically in Fig. 2.12a. More accurate values can be obtained from books of standard mathematical functions. Note that since $\text{erf}(0.5) \approx 0.5$ the depth at which the carbon concentration is midway between C_s and C_0 is given by $(x/2\sqrt{Dt}) \approx 0.5$, that is

$$x \approx \sqrt{Dt} \quad (2.24)$$

Thus the thickness of the carburized layer is $\sim \sqrt{Dt}$. Note also that the depth of any isoconcentration line is directly proportional to $\sqrt{Dt}$, i.e. to obtain a twofold increase in penetration requires a fourfold increase in time.

For the case of carbon diffusion in austenite at 1000 °C, $D \approx 4 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$, which means that a carburized layer 0.2 mm thick requires a time of $(0.2 \times 10^{-3})^2 / 4 \times 10^{-11}$, i.e. 1000 s (17 min).

There are other situations in which the solution to the diffusion equation is very similar to Equation 2.23. For example during decarburization of steel the surface concentration is reduced to a very low value and carbon diffuses

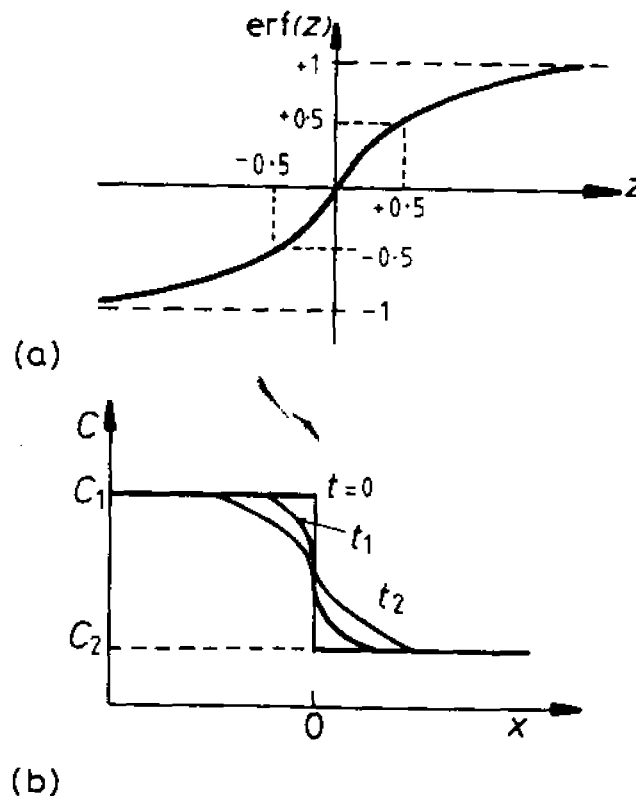


Fig. 2.12 (a) Schematic diagram illustrating the main features of the error function. (b) Concentration profiles at successive times ($t_2 > t_1 > 0$) when two semi-infinite bars of different composition are annealed after welding.

out of the specimen. The carbon profile is then given by

$$C = C_0 \text{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (2.25)$$

Another situation arises if two semi-infinite specimens of different compositions C_1 and C_2 are joined together and annealed. The profiles in this case are shown in Fig. 2.12b and the relevant solution is

$$C = \left(\frac{C_1 + C_2}{2} \right) - \left(\frac{C_1 - C_2}{2} \right) \text{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (2.26)$$

2.3 Substitutional Diffusion

Diffusion in dilute interstitial alloys was relatively simple because the diffusing atoms are always surrounded by 'vacant' sites to which they can jump whenever they have enough energy to overcome the energy barrier for migration. In substitutional diffusion, however, an atom can only jump if there happens to be a vacant site at one of the adjacent lattice positions as shown in Fig. 2.2. The simplest case of substitutional diffusion is the self-diffusion of atoms in a pure metal. This is amenable to a simple atomic model similar to the case of interstitial diffusion and will be treated first. Substitutional diffusion in binary alloys is more complex and will be dealt with separately.

2.3.1 Self-Diffusion

The rate of self-diffusion can be measured experimentally by introducing a few radioactive A atoms (A^*) into pure A and measuring the rate at which penetration occurs at various temperatures. Since A^* and A atoms are *chemically* identical their jump frequencies are also almost identical. Thus the diffusion coefficient can be related to the jump frequency by Equation 2.3, that is

$$D_A^* = D_A = \frac{1}{6} \alpha^2 \Gamma \quad (2.27)$$

where Γ is the jump frequency of both the A^* and A atoms. Strictly speaking, Equation 2.3 was derived on the assumption that each atomic jump is unrelated to the previous jump. This is a good assumption for interstitial diffusion, but it is less valid for substitutional diffusion. The difference is that once an atom has jumped into a vacancy the next jump is not equally probable in all directions, but is most likely to occur back into the same vacancy. Such jumps do not contribute to the diffusive flux and therefore Equation 2.27 should be replaced by $D_A^* = f D_A = f \cdot \alpha^2 \Gamma / 6$ where f (known as a correlation factor) is less than unity. However, the effect is small and f is

close to unity. (See P.G. Shewmon *Diffusion in Solids* McGraw-Hill, New York, 1963, p. 100.)

Consider the atomic jump shown in Fig. 2.2. An atom next to a vacancy can make a jump provided it has enough thermal energy to overcome the activation energy barrier to migration, ΔG_m . Therefore the probability that any attempt at jumping will be successful is given by $\exp(-\Delta G_m/RT)$ as in the case of interstitial migration. However, most of the time the adjacent site will not be vacant and the jump will not be possible. The probability that an adjacent site is vacant is given by zX_v where z is the number of nearest neighbours and X_v is the probability that any one site is vacant, which is just the mole fraction of vacancies in the metal. Combining all these probabilities gives the probability of a successful jump as $zX_v \exp(-\Delta G_m/RT)$. Since the atoms are vibrating with a temperature-independent frequency ν the number of successful jumps any given atom will make in 1 s is given by

$$\Gamma = \nu z X_v \exp \frac{-\Delta G_m}{RT} \quad (2.28)$$

But, if the vacancies are in thermodynamic equilibrium, $X_v = X_v^e$ as given by Equation 1.57, i.e.

$$X_v^e = \exp \frac{-\Delta G_v}{RT} \quad (2.29)$$

Combining these last three equations gives

$$D_A = \frac{1}{6} \alpha^2 \nu \exp \frac{-(\Delta G_m + \Delta G_v)}{RT} \quad (2.30)$$

Substituting $\Delta G = \Delta H - T\Delta S$ gives

$$D_A = \frac{1}{6} \alpha^2 \nu \exp \frac{\Delta S_m + \Delta S_v}{R} \exp - \left(\frac{\Delta H_m + \Delta H_v}{RT} \right) \quad (2.31)$$

For most metals ν is $\sim 10^{13}$. In fcc metals $z = 12$ and $\alpha = a/\sqrt{2}$ the jump distance. This equation can be written more concisely as

$$D_A = D_0 \exp \frac{-Q_{SD}}{RT} \quad (2.32)$$

where

$$D_0 = \frac{1}{6} \alpha^2 \nu \exp \frac{\Delta S_m + \Delta S_v}{R} \quad (2.33)$$

and

$$Q_{SD} = \Delta H_m + \Delta H_v \quad (2.34)$$

Equation 2.32 is the same as was obtained for interstitial diffusion except that

the activation energy for self-diffusion has an extra term (ΔH_v). This is because self-diffusion requires the presence of vacancies whose concentration depends on ΔH_v .

Some of the experimental data on substitutional self-diffusion are summarised in Table 2.2. It can be seen that for a given crystal structure and bond type Q/RT_m is roughly constant; that is, the activation enthalpy for self-diffusion, Q , is roughly proportional to the equilibrium melting temperature, T_m . Also, within each class, the diffusivity at the melting temperature, $D(T_m)$, and D_0 are approximately constants. For example, for most close-packed metals (fcc and hcp) $Q/RT_m \sim 18$ and $D(T_m) \sim 1 \mu\text{m}^2 \text{s}^{-1}$ ($10^{-12} \text{m}^2 \text{s}^{-1}$). The Q/RT_m and $D(T_m)$ data are also plotted in Fig. 2.13 along with data for other materials for comparison. An immediate consequence of these correlations is that the diffusion coefficients of all materials with a given crystal structure and bond type will be approximately the same at the same fraction of their melting temperature, i.e. $D(T/T_m) = \text{constant}$. (T/T_m is known as the homologous temperature.)

The above correlations have been evaluated for atmospheric pressure. There are, however, limited experimental data that suggest the same correlations hold independently of pressure, provided of course the effect of pressure on T_m is taken into account. Since volume usually increases on melting, raising the pressure increases T_m and thereby lowers the diffusivity at a given temperature.

That a rough correlation exists between Q and T_m is not surprising: increasing the interatomic bond strength makes the process of melting more difficult; that is, T_m is raised. It also makes diffusion more difficult by increasing ΔH_v and ΔH_m .

Consider the effect of temperature on self-diffusion in Cu as an example. At 800 °C (1073 K) the data in Table 2.2 give $D_{Cu} = 5 \times 10^{-9} \text{mm}^2 \text{s}^{-1}$. The jump distance α in Cu is 0.25 nm and Equation 2.3 therefore gives $\Gamma_{Cu} = 5 \times 10^5 \text{jumps s}^{-1}$. After an hour at this temperature, $\sqrt{(Dt)} \sim 4 \mu\text{m}$. Extrapolating the data to 20 °C, however, gives $D_{Cu} \sim 10^{-34} \text{mm}^2 \text{s}^{-1}$, i.e. $\Gamma \sim 10^{-20} \text{jumps s}^{-1}$. Alternatively, each atom would make one jump every 10^{12} years!

Experimentally the usual method for determining the self-diffusion coefficient is to deposit a known quantity (M) of a radioactive isotope A^* onto the ends of two bars of A which are then joined as shown in Fig. 2.14a. After annealing for a known time at a fixed temperature, A^* will have diffused into A and the concentration profile can be determined by machining away thin layers of the bar and measuring the radioactivity as a function of position. Since A and A^* are chemically identical the diffusion of A^* into A will occur according to Equation 2.18. The solution of this equation for the present boundary conditions is

$$C = \frac{M}{2\sqrt{(\pi Dt)}} \exp \left(-\frac{x^2}{4Dt} \right) \quad (2.35)$$

Table 2.2 Experimental Data for Substitutional Self-Diffusion in Pure Metals at Atmospheric Pressure

Data selected mainly from A.M. Brown and M.F. Ashby, 'Correlations for Diffusion Constants', *Acta Metallurgica*, **28**: 1085 (1980).

Class	Metal	T_m K	D_0 $\text{mm}^2 \text{s}^{-1}$	Q kJ mol^{-1}	Q RT_m	$D(T_m)$ $\mu\text{m}^2 \text{s}^{-1}$
bcc (rare earths)	ϵ -Pu	914	0.3	65.7	8.7	53
	δ -Ce	1071	1.2	90.0	10.1	49
	γ -La	1193	1.3	102.6	10.4	42
	γ -Yb	1796	1.2	121.0	8.1	3600
bcc (alkali metals)	Rb	312	23	39.4	15.2	5.8
	K	337	31	40.8	14.6	15
	Na	371	24.2	43.8	14.2	16
	Li	454	23	55.3	14.7	9.9
bcc (transition metals)	β -Ti	577	40	94.6	19.7	0.11
	Eu	1095	100	143.5	15.8	14
	Er	1795	451	302.4	20.3	0.71
	α -Fe*	1811	200	239.7	15.9	26
	δ -Fe*	1811	190	238.5	15.8	26
	β -Ti	1933	109	251.2	15.6	18
	β -Zr	2125	134	273.5	15.5	25
	Cr	2130	20	308.6	17.4	0.54
	V	2163	28.8	309.2	17.2	0.97
	Nb	2741	1240	439.6	19.3	5.2
	Mo	2890	180	460.6	19.2	0.84
	Ta	3269	124	413.3	15.2	31
	W	3683	4280	641.0	20.9	3.4
hcp*	Cd	594	$\parallel c$ 5	76.2	15.4	0.99
			$\perp c$ 10	79.9	16.2	0.94
	Zn	692	$\parallel c$ 13	91.6	15.9	1.6
			$\perp c$ 18	96.2	16.7	0.98
	Mg	922	$\parallel c$ 100	134.7	17.6	2.3
			$\perp c$ 150	136.0	17.8	2.9
fcc	Pb	601	137	109.1	21.8	0.045
	Al	933	170	142.0	18.3	1.9
	Ag	1234	40	184.6	18.0	0.61
	Au	1336	10.7	176.9	15.9	1.3
	Cu	1356	31	200.3	17.8	0.59
	Ni	1726	190	279.7	19.5	0.65

Table 2.2 (cont.)

Class	Metal	T_m K	D_0 $\text{mm}^2 \text{s}^{-1}$	Q kJ mol^{-1}	Q RT_m	$D(T_m)$ $\mu\text{m}^2 \text{s}^{-1}$
	β -Co	1768	83	283.4	19.3	0.35
	γ -Fe†	1805	49	284.1	18.9	0.29
	Pd	1825	20.5	266.3	17.6	0.49
	Th	2023	120	319.7	19.0	6.6
	Pt	2046	22	278.4	16.4	0.17
tet*	β -Sn	505	$\parallel c$ 770	107.1	25.5	0.0064
			$\perp c$ 1070	105.0	25.0	0.015
diamond cubic	Ge	1211	440	324.5	32.3	4.4×10^{-5}
	Si	1683	0.9×10^6	496.0	35.5	3.6×10^{-4}

* Data selected from N.L. Peterson, *Solid State Physics*, Vol. 22, D. Turnbull and H. Ehrenreich (Eds.), Academic Press, New York, 1968.

† T_m for γ -Fe is the temperature at which γ -Fe would melt if δ -Fe did not intervene.

M has units [quantity m^{-2}] and C [quantity m^{-3}]. Figure 2.14b shows the form of this equation fitted to experimental points for self-diffusion in gold.

2.3.2 Vacancy Diffusion

The jumping of atoms into vacant sites can equally well be considered as the jumping of vacancies onto atom sites. If excess vacancies are introduced into the lattice they will diffuse at a rate which depends on the jump frequency. However, a vacancy is always surrounded by sites to which it can jump and it is thus analogous to an interstitial atom (see Section 2.2.2). Therefore a vacancy can be considered to have its own diffusion coefficient given by

$$D_v = \frac{1}{6} \alpha^2 \Gamma_v \quad (2.36)$$

By analogy with Equation 2.8

$$D_v = \frac{1}{6} \alpha^2 z v \exp \frac{\Delta S_m}{R} \exp \frac{-\Delta H_m}{RT} \quad (2.37)$$

In this case ΔH_m and ΔS_m apply to the migration of a vacancy, and are therefore the same as for the migration of a substitutional atom. Comparing Equations 2.37 and 2.31 it can be seen that

$$D_v = D_A / X_v^c \quad (2.38)$$

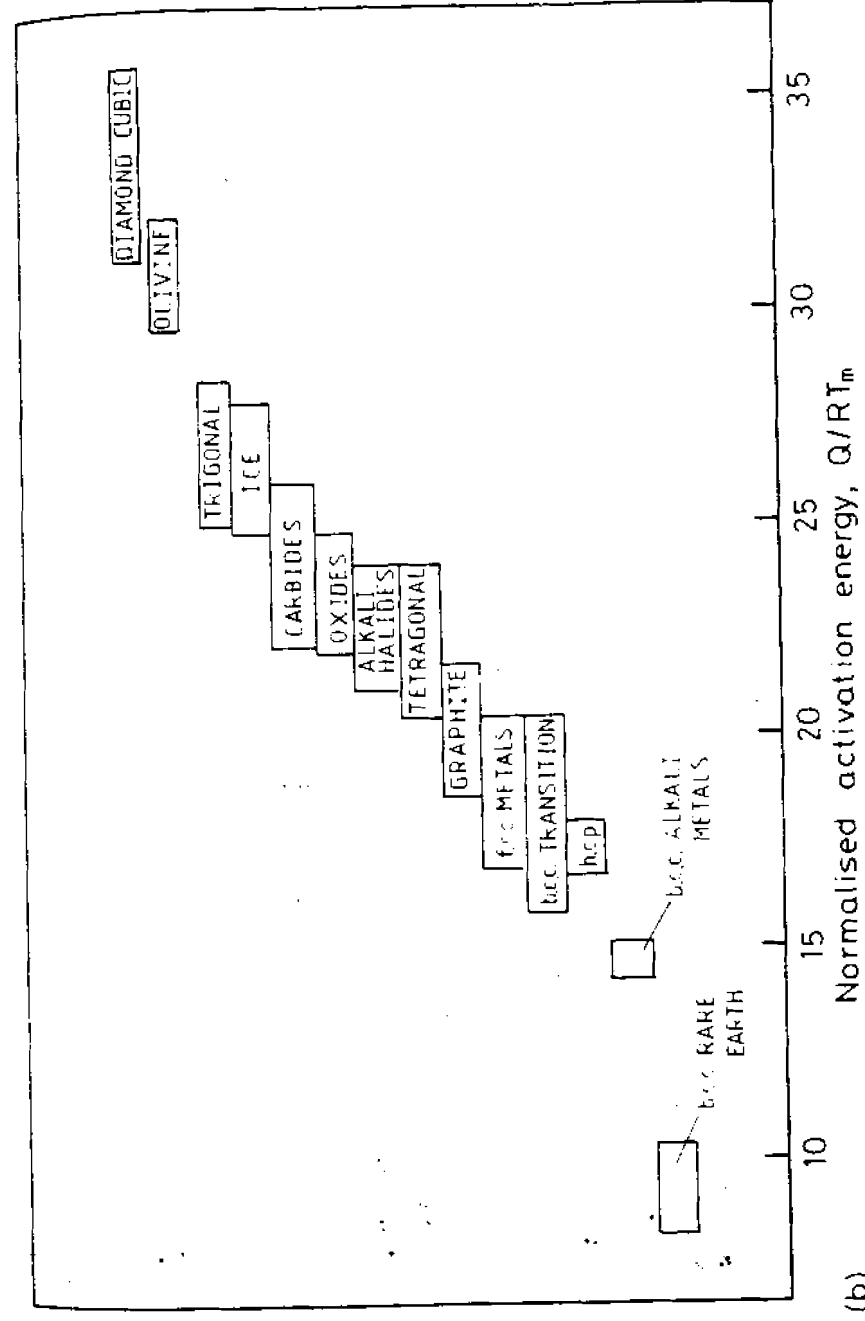
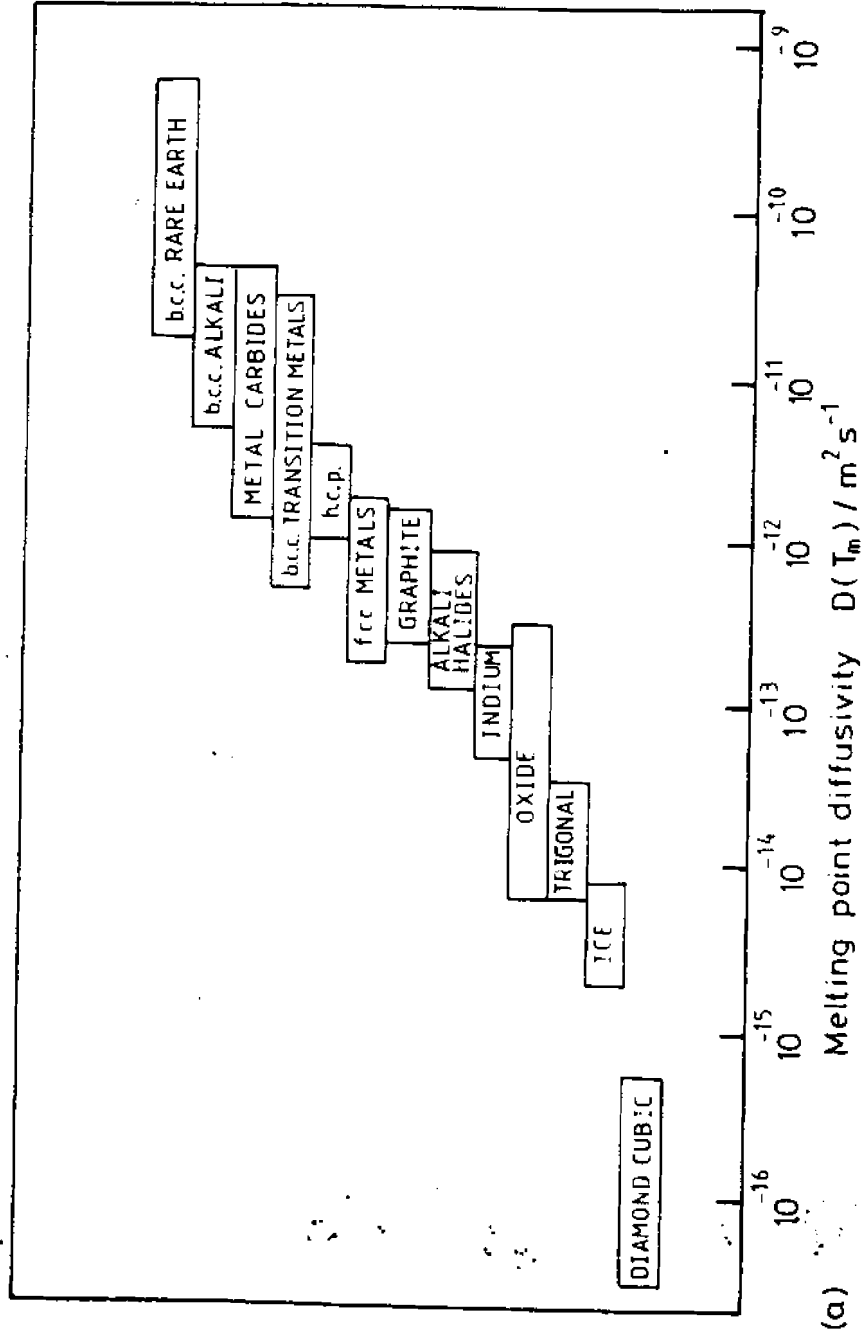


Fig. 2.13 (a) Melting point diffusivities and (b) normalised activation energies for various classes of materials. (After A.M. Brown and M.F. Ashby, *Acta Metallurgica* 28 (1980) 1085.)

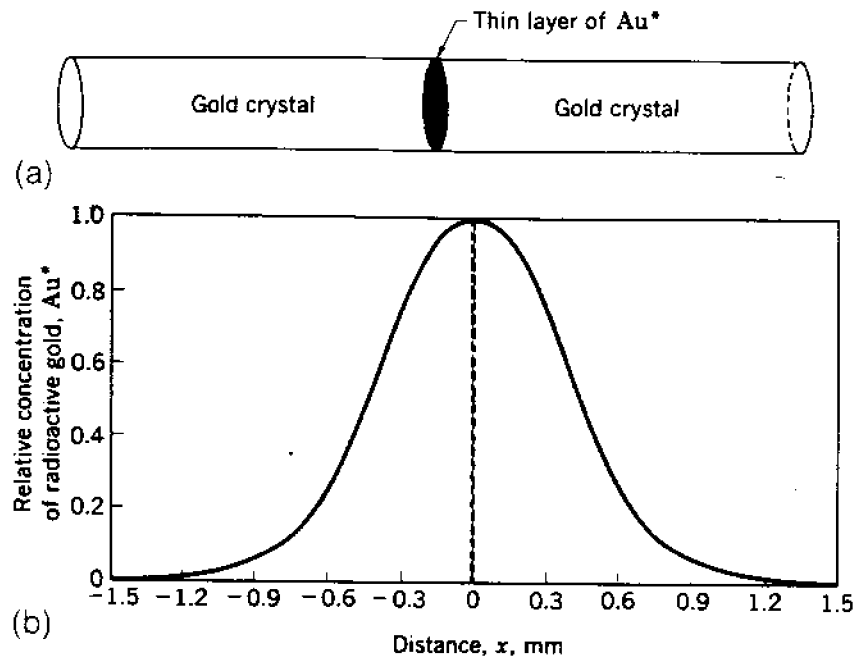


Fig. 2.14 Illustration of the principle of tracer diffusion and of the planar source method for determining the self-diffusion coefficient of gold. (a) Initial diffusion couple with planar source of radioactive gold Au*. (b) Distribution of Au* after diffusion for 100 h at 920 °C. (After A.G. Guy, *Introduction to Materials Science*, McGraw-Hill, New York, 1971.)

This shows in fact that D_v is many orders of magnitude greater than D_A the diffusivity of substitutional atoms.

2.3.3 Diffusion in Substitutional Alloys

During self-diffusion all atoms are chemically identical. Thus the probability of finding a vacancy adjacent to any atom and the probability that the atom will make a jump into the vacancy is equal for all atoms. This leads to a simple relationship between jump frequency and diffusion coefficient. In binary substitutional alloys, however, the situation is more complex. In general, the rate at which solvent (A) and solute (B) atoms can move into a vacant site is not equal and each atomic species must be given its own *intrinsic* diffusion coefficient D_A or D_B .

The fact that the A and B atoms occupy the same sites has important consequences on the form that Fick's first and second laws assume for substitutional alloys. It will be seen later that when the A and B atoms jump at different rates the presence of concentration gradients induces a movement of the lattice through which the A and B atoms are diffusing.

D_A and D_B are defined such that Fick's first law applies to diffusion *relative to the lattice*, that is

$$J_A = -D_A \frac{\partial C_A}{\partial x} \quad (2.39)$$

$$J_B = -D_B \frac{\partial C_B}{\partial x} \quad (2.40)$$

where J_A and J_B are the fluxes of A and B atoms *across a given lattice plane*. This point did not need emphasizing in the case of interstitial diffusion because the lattice planes of the parent atoms were unaffected by the diffusion process. It will be seen, however, that the situation is different in the case of substitutional diffusion.

In order to derive Fick's second law let us consider the interdiffusion of A and B atoms in a diffusion couple that is made by welding together blocks of pure A and B as shown in Fig. 2.15a. If the couple is annealed at a high enough temperature, a concentration profile will develop as shown.

If we make the simplifying assumption that the total number of atoms per unit volume is a constant, C_0 , independent of composition, then

$$C_0 = C_A + C_B \quad (2.41)$$

and

$$\frac{\partial C_A}{\partial x} = -\frac{\partial C_B}{\partial x} \quad (2.42)$$

Hence at a given position the concentration gradients driving the diffusion of A and B atoms are equal but opposite, and the fluxes of A and B *relative to the lattice* can be written as

$$J_A = -D_A \frac{\partial C_A}{\partial x} \quad (2.43)$$

$$J_B = D_B \frac{\partial C_A}{\partial x}$$

These fluxes are shown schematically in Fig. 2.15 for the case $D_A > D_B$, i.e. $|J_A| > |J_B|$.

When atoms migrate by the vacancy process the jumping of an atom into a vacant site can equally well be regarded as the jumping of the vacancy onto the atom, as illustrated in Fig. 2.16. In other words, if there is a net flux of atoms in one direction there is an equal flux of vacancies in the opposite direction. Thus in Fig. 2.15a there is a flux of vacancies $-J_A$ due to the migration of A atoms *plus* a flux of vacancies $-J_B$ due to the diffusion of B atoms. As $J_A > J_B$ there will be a net flux of vacancies

$$J_v = -J_A - J_B \quad (2.44)$$

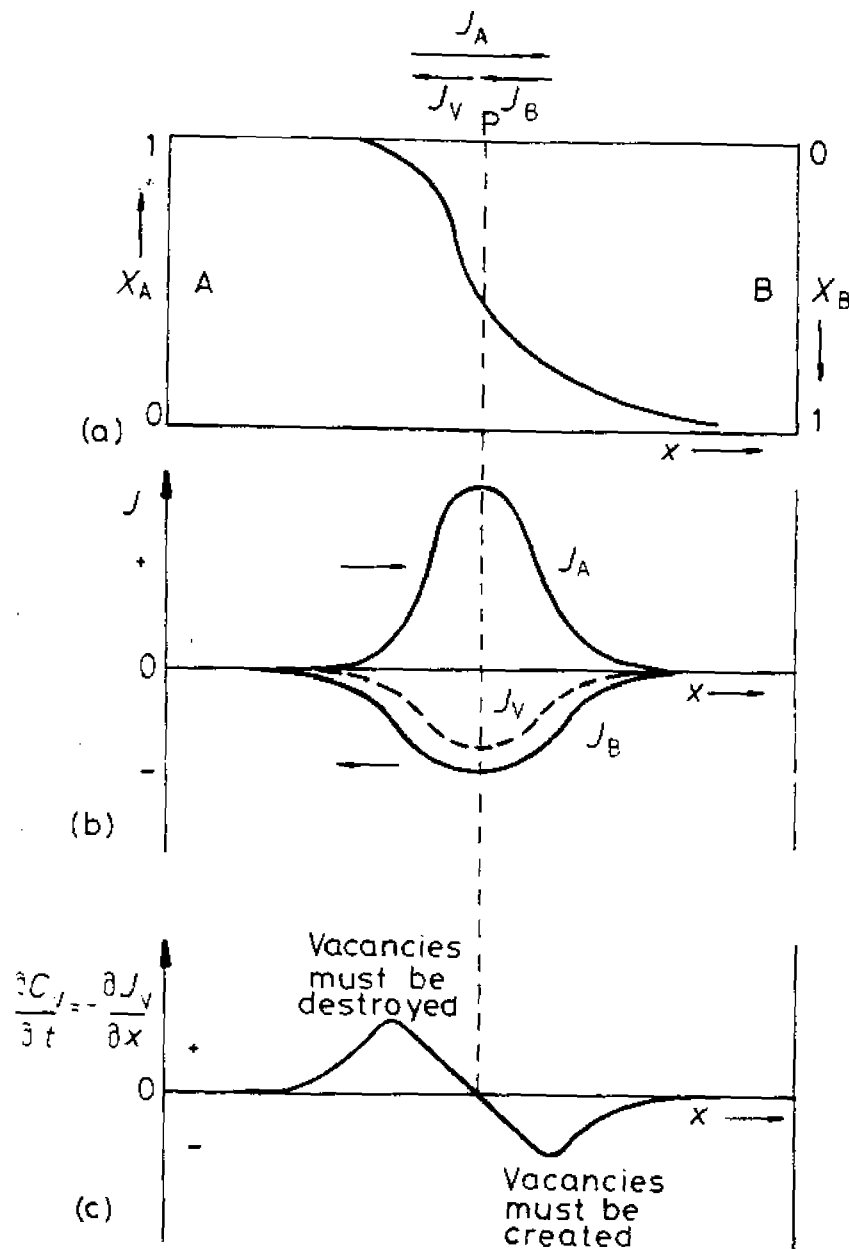


Fig. 2.15 Interdiffusion and vacancy flow. (a) Composition profile after interdiffusion of A and B. (b) The corresponding fluxes of atoms and vacancies as a function of position x . (c) The rate at which the vacancy concentration would increase or decrease if vacancies were not created or destroyed by dislocation climb.

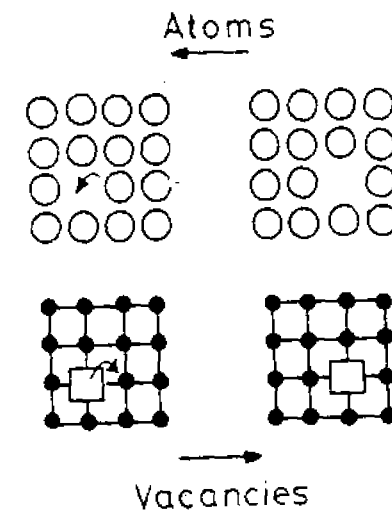


Fig. 2.16 The jumping of atoms in one direction can be considered as the jumping of vacancies in the other direction.

This is indicated in vector notation in Fig. 2.15a. In terms of D_A and D_B , therefore

$$J_v = (D_A - D_B) \frac{\partial C_A}{\partial x} \quad (2.45)$$

This leads to a variation in J_v across the diffusion couple as illustrated in Fig. 2.15b.

In order to maintain the vacancy concentration everywhere near equilibrium vacancies must be created on the B-rich side and destroyed on the A-rich side. The rate at which vacancies are *created* or *destroyed* at any point is given by $\partial C_v / \partial t = -\partial J_v / \partial x$ (Equation 2.16) and this varies across the diffusion couple as shown in Fig. 2.15c.

It is the net flux of vacancies across the middle of the diffusion couple that gives rise to movement of the lattice. Jogged edge dislocations can provide a convenient source or sink for vacancies as shown in Fig. 2.17. Vacancies can be absorbed by the extra half-plane of the edge dislocation shrinking while growth of the plane can occur by the emission of vacancies. If this or a similar mechanism operates on each side of the diffusion couple then the required flux of vacancies can be generated as illustrated in Fig. 2.18. This means that extra atomic planes will be introduced on the B-rich side while whole planes of atoms will be 'eaten' away on the A-rich side. Consequently the lattice planes in the middle of the couple will be shifted to the left.

The velocity at which any given lattice plane moves, v , can be related to the flux of vacancies crossing it. If the plane has an area A , during a small time interval δt , the plane will sweep out a volume of $A v \cdot \delta t$ containing $A v \cdot \delta t \cdot C_0$ atoms. This number of atoms is removed by the total number of vacancies crossing the plane in the same time interval, i.e. $J_v A \cdot \delta t$, giving

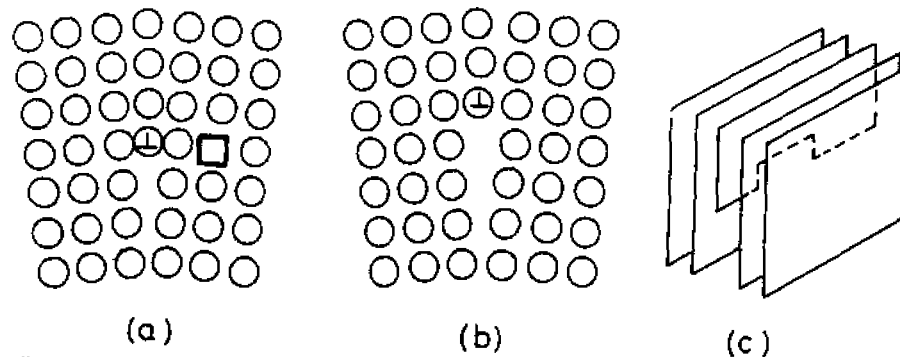


Fig. 2.17 (a) before, (b) after: a vacancy is absorbed at a jog on an edge dislocation (positive climb). (b) before, (a) after: a vacancy is created by negative climb of an edge dislocation. (c) Perspective drawing of a jogged edge dislocation.

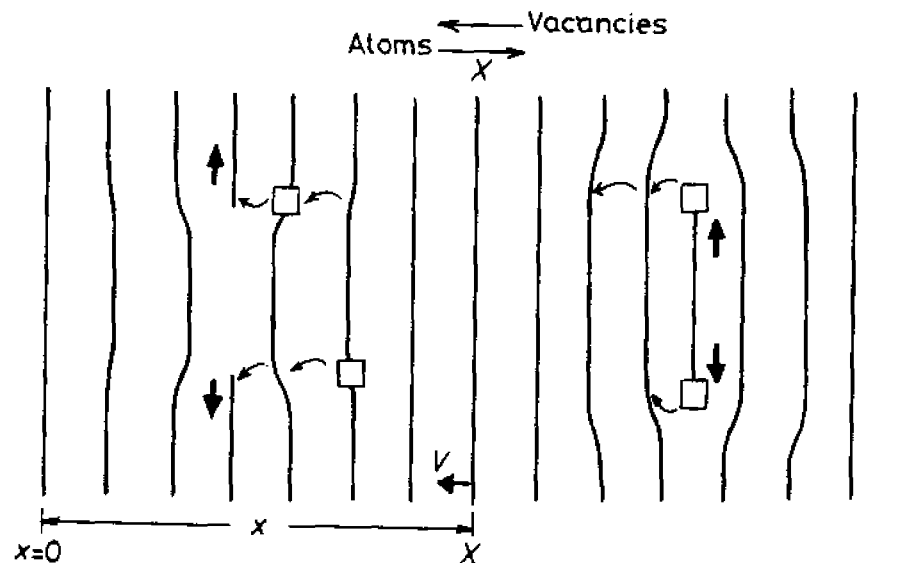


Fig. 2.18 A flux of vacancies causes the atomic planes to move through the specimen.

$$J_v = C_0 v \quad (2.46)$$

Thus the velocity of the lattice planes will vary across the couple in the same way as J_v , see Fig. 2.15b. Substituting Equation 2.45 gives

$$v = (D_A - D_B) \frac{\partial X_A}{\partial x} \quad (2.47)$$

where the mole fraction of A, $X_A = C_A/C_0$

In practice, of course, internal movements of lattice planes are usually not directly of interest. More practical questions concern how long homogenization of an alloy takes, or how rapidly the composition will change at a fixed

position relative to the ends of a specimen. To answer these questions we can derive Fick's second law for substitutional alloys.

Consider a thin slice of material δx thick at a fixed distance x from one end of the couple which is outside the diffusion zone as shown in Fig. 2.19. If the total flux of A atoms entering this slice across plane 1 is J'_A and the total flux leaving is $J'_A + (\partial J'_A / \partial x) \delta x$ the same arguments as were used to derive Equation 2.16 can be used to show that

$$\frac{\partial C_A}{\partial t} = - \frac{\partial J'_A}{\partial x} \quad (2.48)$$

The total flux of A atoms across a stationary plane with respect to the specimen is the sum of two contributions: (i) a diffusive flux $J_A = -D_A \partial C_A / \partial x$ due to diffusion relative to the lattice, and (ii) a flux $v \cdot C_A$ due to the velocity of the lattice in which diffusion is occurring. Therefore:

$$J'_A = -D_A \frac{\partial C_A}{\partial x} + v C_A \quad (2.49)$$

By combining this equation with Equation 2.47 we obtain the equivalent of Fick's first law for the flux relative to the specimen ends:

$$J'_A = -(X_B D_A + X_A D_B) \frac{\partial C_A}{\partial x} \quad (2.50)$$

where $X_A = C_A/C_0$ and $X_B = C_B/C_0$ are the mole fractions of A and B respectively. This can be simplified by defining an interdiffusion coefficient $\bar{D}$ as

$$\bar{D} = X_B D_A + X_A D_B \quad (2.51)$$

so that Fick's first law becomes

$$J'_A = -\bar{D} \frac{\partial C_A}{\partial x} \quad (2.52)$$

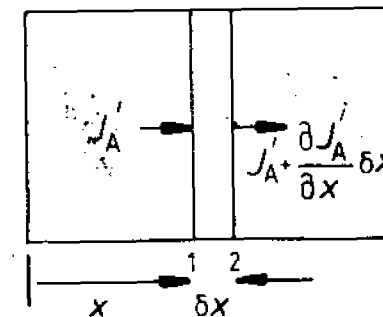


Fig. 2.19 Derivation of Fick's second law for interdiffusion. (See text for details.)

Likewise,

$$J'_B = -\bar{D} \frac{\partial C_B}{\partial x} = \bar{D} \frac{\partial C_A}{\partial x}$$

i.e.

$$J'_B = -J'_A$$

Substitution of Equation 2.52 into Equation 2.48 gives

$$\frac{\partial C_A}{\partial t} = \frac{\partial}{\partial x} \left(\bar{D} \frac{\partial C_A}{\partial x} \right) \quad (2.53)$$

This equation is Fick's second law for diffusion in substitutional alloys. The only difference between this equation and Equation 2.18 (for interstitial diffusion) is that the *interdiffusion* coefficient $\bar{D}$ for substitutional alloys depends on D_A and D_B whereas in interstitial diffusion D_B alone is needed. Equations 2.47 and 2.51 were first derived by Darken⁵ and are usually known as *Darken's equations*.

By solving Equation 2.53 with appropriate boundary conditions it is possible to obtain $C_A(x, t)$ and $C_B(x, t)$, i.e. the concentration of A and B at any position (x) after any given annealing time (t). The solutions that were given in Section 2.2.5 will be applicable to substitutional alloys provided the range of compositions is small enough that any effect of composition on $\bar{D}$ can be ignored. For example, if $\bar{D}$ is known the characteristic relaxation time for an homogenization anneal would be given by Equation 2.21 using $\bar{D}$ in place of D_B , i.e.

$$\tau = \frac{l^2}{\pi^2 \bar{D}} \quad (2.54)$$

If the initial composition differences are so great that changes in $\bar{D}$ become important then more complex solutions to Equation 2.53 must be used. These will not be dealt with here, however, as they only add mathematical complexities without increasing our understanding of the basic principles⁶.

Experimentally it is possible to measure $\bar{D}$ by determining the variation of X_A or X_B after annealing a diffusion couple for a given time such as that shown in Fig. 2.15a. In cases where $\bar{D}$ can be assumed constant a comparison of Equation 2.26 and the measured concentration profile would give $\bar{D}$. When $\bar{D}$ is not constant there are graphical solutions to Fick's second law that enable $\bar{D}$ to be determined at any composition. In order to determine D_A and D_B separately it is also necessary to measure the velocity of the lattice at a given point in the couple. This can be achieved in practice by inserting insoluble wires at the interface before welding the two blocks together. These wires remain in effect 'fixed' to the lattice planes and their displacement after a given annealing time can be used to calculate v . When v and $\bar{D}$ are known, Equations 2.47 and 2.51 can be used to calculate D_A and D_B for the composition at the markers.

The displacement of inert wires during diffusion was first observed by

Smigelskas and Kirkendall in 1947⁷ and is usually known as the *Kirkendall effect*. In this experiment a block of α -brass (Cu-30wt% Zn) was wound with molybdenum wire and encapsulated in a block of pure Cu, as shown in Fig. 2.20. After annealing at a high temperature it was found that the separation of the markers (w) had decreased. This is because $D_{Zn} > D_{Cu}$ and the zinc atoms diffuse out of the central block faster than they are replaced by copper atoms diffusing in the opposite direction. Similar effects have since been demonstrated in many other alloy systems. In general it is found that in any given couple, atoms with the lower melting point possess a higher D . The exact value of D , however, varies with the composition of the alloy. Thus in Cu-Ni alloys D_{Cu} , D_{Ni} and $\bar{D}$ are all composition dependent, increasing as X_{Cu} increases. Fig. 2.21.

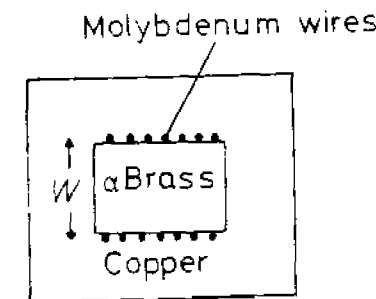


Fig. 2.20 An experimental arrangement to show the Kirkendall effect.

In Fig. 2.17 it was assumed that the extra half-planes of atoms that grew or shrank due to the addition or loss of atoms, were parallel to the original weld interface so that there were no constraints on the resultant local expansion or contraction of the lattice. In practice, however, these planes can be oriented in many directions and the lattice will also try to expand or contract parallel to the weld interface. Such volume changes are restricted by the surrounding material with the result that two-dimensional compressive stresses develop in regions where vacancies are created, while tensile stresses arise in regions where vacancies are destroyed. These stress fields can even induce plastic deformation resulting in microstructures characteristic of hot deformation.

Vacancies are not necessarily all annihilated at dislocations, but can also be absorbed by internal boundaries and free surfaces. However, those not absorbed at dislocations mainly agglomerate to form holes or voids in the lattice. Void nucleation is difficult because it requires the creation of a new surface and it is generally believed that voids are heterogeneously nucleated at impurity particles. The tensile stresses that arise in conjunction with vacancy destruction can also play a role in the nucleation of voids. When voids are formed the equations derived above cannot be used without modification.

In concentrated alloys the experimentally determined values of $\bar{D}$, D_A and D_B are also found to show the same form of temperature dependence as all

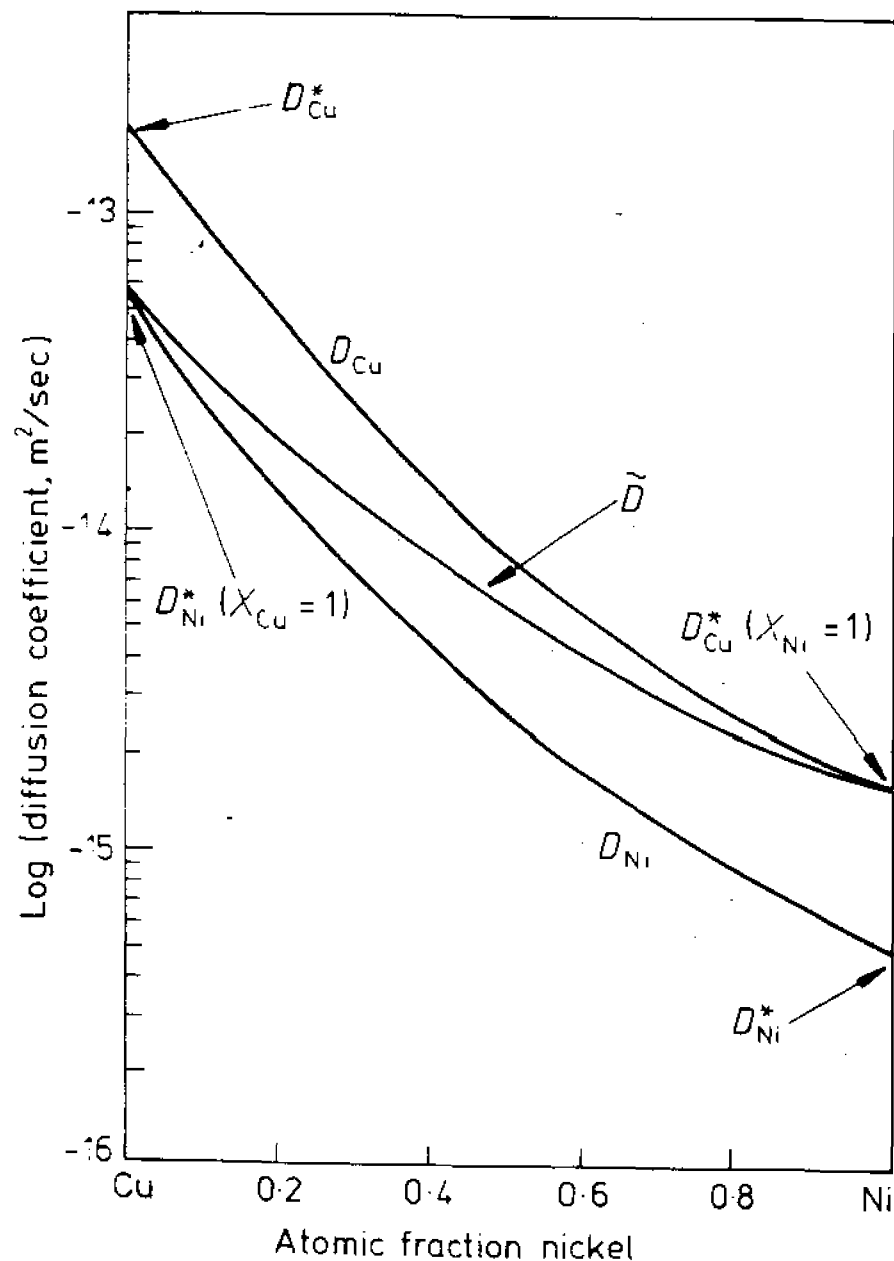


Fig. 2.21 The relationship between the various diffusion coefficients in the Cu-Ni system at 1000 °C (After A.G. Guy, *Introduction to Materials Science*, McGraw-Hill, New York, 1971.)

other diffusivities, so that

$$\tilde{D} = \tilde{D}_0 \exp \frac{-Q}{RT} \quad (2.55)$$

$$D_A = D_{A0} \exp \frac{-Q_A}{RT} \quad (2.56)$$

$$D_B = D_{B0} \exp \frac{-Q_B}{RT} \quad (2.57)$$

However the factors that determine D_0 and Q in these cases are uncertain and there is no simple atomistic model for concentrated solutions.

The variation of $\tilde{D}$ with composition can be estimated in cases where it has not been measured, by utilizing two experimental observations⁸:

1. For a given crystal structure, $\tilde{D}$ at the melting point is roughly constant. Therefore if adding B to A decreases the melting point, $\tilde{D}$ will increase, at a given temperature, and vice versa.
2. For a given solvent and temperature, both interstitial and substitutional diffusion are more rapid in a bcc lattice than a close-packed lattice. For example, for the diffusion of carbon in Fe at 910 °C, $D_C^{\alpha}/D_C^{\gamma} \sim 100$. At 850 °C the self-diffusion coefficients for Fe are such that $D_{Fe}^{\alpha}/D_{Fe}^{\gamma} \sim 100$. The reason for this difference lies in the fact that the bcc structure is more open and the diffusion processes require less lattice distortion.

2.3.4 Diffusion in Dilute Substitutional Alloys

Another special situation arises with diffusion in dilute alloys. When $X_B \sim 0$ and $X_A \sim 1$, Equation 2.51 becomes

$$\tilde{D} = D_B \quad (2.58)$$

This is reasonable since it means that the rate of homogenization in dilute alloys is controlled by how fast the solute (B) atoms can diffuse. Indeed the only way homogenization can be achieved is by the migration of the B atoms into the solute-depleted regions. D_B for a dilute solution of B in A is called the impurity diffusion coefficient. Such data is more readily available than interdiffusion data in concentrated alloys. One way in which impurity diffusion coefficients can be measured is by using radioactive tracers.

It is often found that D_B in a dilute solution of B in A is greater than D_A . The reason for this is that the solute atoms can attract vacancies so that there is *more* than a random probability of finding a vacancy next to a solute atom with the result that they can diffuse faster than the solvent. An attraction between a solute atom and a vacancy can arise if the solute atom is larger than the solvent atoms or if it has higher valency. If the binding energy is very large the vacancy will be unable to 'escape' from the solute atom. In this case the solute-vacancy pair can diffuse through the lattice together.

2.4 Atomic Mobility

Fick's first law is based on the assumption that diffusion eventually stops, that is equilibrium is reached, when the concentration is the same everywhere. Strictly speaking this situation is never true in practice because real materials always contain lattice defects such as grain boundaries, phase boundaries and dislocations. Some atoms can lower their free energies if they migrate to such defects and at 'equilibrium' their concentrations will be higher in the vicinity of the defect than in the matrix. Diffusion in the vicinity of these defects is therefore affected by both the concentration gradient and the gradient of the interaction energy. Fick's law alone is insufficient to describe how the concentration will vary with distance and time.

As an example consider the case of a solute atom that is too big or too small in comparison to the space available in the solvent lattice. The potential energy of the atom will then be relatively high due to the strain in the surrounding matrix. However, this strain energy can be reduced if the atom is located in a position where it better matches the space available, e.g. near dislocations and in boundaries, where the matrix is already distorted.

Segregation of atoms to grain boundaries, interfaces and dislocations is of great technological importance. For example the diffusion of carbon or nitrogen to dislocations in mild steel is responsible for strain ageing and blue brittleness. The segregation of impurities such as Sb, Sn, P and As to grain boundaries in low-alloy steels produces temper embrittlement. Segregation to grain boundaries affects the mobility of the boundary and has pronounced effects on recrystallization, texture and grain growth. Similarly the rate at which phase transformations occur is sensitive to segregation at dislocations and interfaces.

The problem of atom migration can be solved by considering the thermodynamic condition for equilibrium; namely that the chemical potential of an atom must be the same everywhere. Diffusion continues in fact until this condition is satisfied. Therefore it seems reasonable to suppose that in general the flux of atoms at any point in the lattice is proportional to the chemical potential gradient. Fick's first law is merely a special case of this more general approach.

An alternative way to describe a flux of atoms is to consider a net drift velocity (v) superimposed on the random jumping motion of each diffusing atom. The drift velocity is simply related to the diffusive flux via the equation

$$J_B = v_B C_B \quad (2.59)$$

Since atoms always migrate so as to remove differences in chemical potential it is reasonable to suppose that the drift velocity is proportional to the local chemical potential gradient, i.e.

$$v_B = -M_B \frac{\partial \mu_B}{\partial x} \quad (2.60)$$

where M_B is a constant of proportionality known as the atomic mobility. Since μ_B has units of energy the derivative of μ_B with respect to distance ($\partial \mu_B / \partial x$) is effectively the chemical 'force' causing the atom to migrate.

Combining Equations 2.59 and 2.60 gives

$$J_B = -M_B C_B \frac{\partial \mu_B}{\partial x} \quad (2.61)$$

Intuitively it seems that the mobility of an atom and its diffusion coefficient must be closely related. The relationship can be obtained by relating $\partial \mu / \partial x$ to $\partial C / \partial x$ for a stress-free solid solution. Using Equation 1.70 and $C_B = X_B / V_m$ Equation 2.61 becomes

$$J_B = -M_B \frac{X_B}{V_m} \cdot \frac{RT}{X_B} \left\{ 1 + \frac{d \ln \gamma_B}{d \ln X_B} \right\} \frac{\partial X_B}{\partial x} \quad (2.62)$$

i.e.

$$J_B = -M_B RT \left\{ 1 + \frac{d \ln \gamma_B}{d \ln X_B} \right\} \frac{\partial C_B}{\partial x} \quad (2.63)$$

Comparison with Fick's first law gives the required relationship:

$$D_B = M_B RT \left\{ 1 + \frac{d \ln \gamma_B}{d \ln X_B} \right\} \quad (2.64)$$

Similarly

$$D_A = M_A RT \left\{ 1 + \frac{d \ln \gamma_A}{d \ln X_A} \right\} \quad (2.65)$$

For *ideal* or *dilute* solutions ($X_B \rightarrow 0$) γ_B is a constant and the term in brackets is unity, i.e.

$$D_B = M_B RT \quad (2.66)$$

For non-ideal concentrated solutions the terms in brackets, the so-called thermodynamic factor, must be included. As shown by Equation 1.71 this factor is the same for both A and B and is simply related to the curvature of the molar free energy-composition curve.

When diffusion occurs in the presence of a strain energy gradient, for example, the expression for the chemical potential can be modified to include the effect of an elastic strain energy term E which depends on the position (x) relative to a dislocation, say

$$\mu_B = G_B + RT \ln \gamma_B X_B + E \quad (2.67)$$

Following the above procedure, this gives

$$J_B = -D_B \frac{\partial C_B}{\partial x} - \frac{D_B C_B}{RT} \frac{\partial E}{\partial x} \quad (2.68)$$

It can thus be seen that in addition to the effect of the concentration gradient the diffusive flux is also affected by the gradient of strain energy, $\partial E/\partial x$.

Other examples of atoms diffusing towards regions of high concentration can be found when diffusion occurs in the presence of an electric field or a temperature gradient. These are known as electromigration and thermomigration respectively⁹. Cases encountered in phase transformations can be found where atoms migrate across phase boundaries, or, as mentioned in the introduction, when the free energy curve has a negative curvature. The latter is known as spinodal decomposition.

2.5 Tracer Diffusion in Binary Alloys

The use of radioactive tracers were described in connection with self-diffusion in pure metals. It is, however, possible to use radioactive tracers to determine the intrinsic diffusion coefficients of the components in an alloy. The method is similar to that shown in Fig. 2.14 except that a small quantity of a suitable radioactive tracer, e.g. B^* , is allowed to diffuse into a homogeneous bar of A/B solution. The value obtained for D from Equation 2.35 is the tracer diffusion coefficient D_B^* .

Such experiments have been carried out on a whole series of gold-nickel alloys at 900 °C¹⁰. At this temperature gold and nickel are completely soluble in each other, Fig. 2.22a. The results are shown in Fig. 2.22c. Since radioactive isotopes are chemically identical it might appear at first sight that the tracer diffusivities (D_{Au}^* and D_{Ni}^*) should be identical to the intrinsic diffusivities (D_{Au} and D_{Ni}) determined by marker movement in a diffusion couple. This would be convenient as the intrinsic diffusivities are of more practical value whereas it is much easier to determine tracer diffusivities. However, it can be demonstrated that this is not the case. D_{Au}^* gives the rate at which Au^* (or Au) atoms diffuse in a *chemically homogeneous* alloy, whereas D_{Au} gives the diffusion rate of Au when a concentration gradient is present.

The Au-Ni phase diagram contains a miscibility gap at low temperatures implying that $\Delta H_{mix} > 0$ (the gold and nickel atoms 'dislike' each other). Therefore, whereas the jumps made by Au atoms in a chemically homogeneous alloy will be equally probable in all directions, in a concentration gradient they will be biased away from the Ni-rich regions. The rate of homogenization will therefore be slower in the second case, i.e. $D_{Au} < D_{Au}^*$ and $D_{Ni} < D_{Ni}^*$. On the other hand since the chemical potential gradient is the driving force for diffusion in both types of experiment it is reasonable to suppose that the atomic mobilities are not affected by the concentration gradient. If this is true the intrinsic chemical diffusivities and tracer diffusivities can be related as follows.

In the tracer diffusion experiment the tracer essentially forms a dilute solution in the alloy. Therefore from Equation 2.66

$$D_B^* = M_B^* RT = M_B RT \quad (2.69)$$

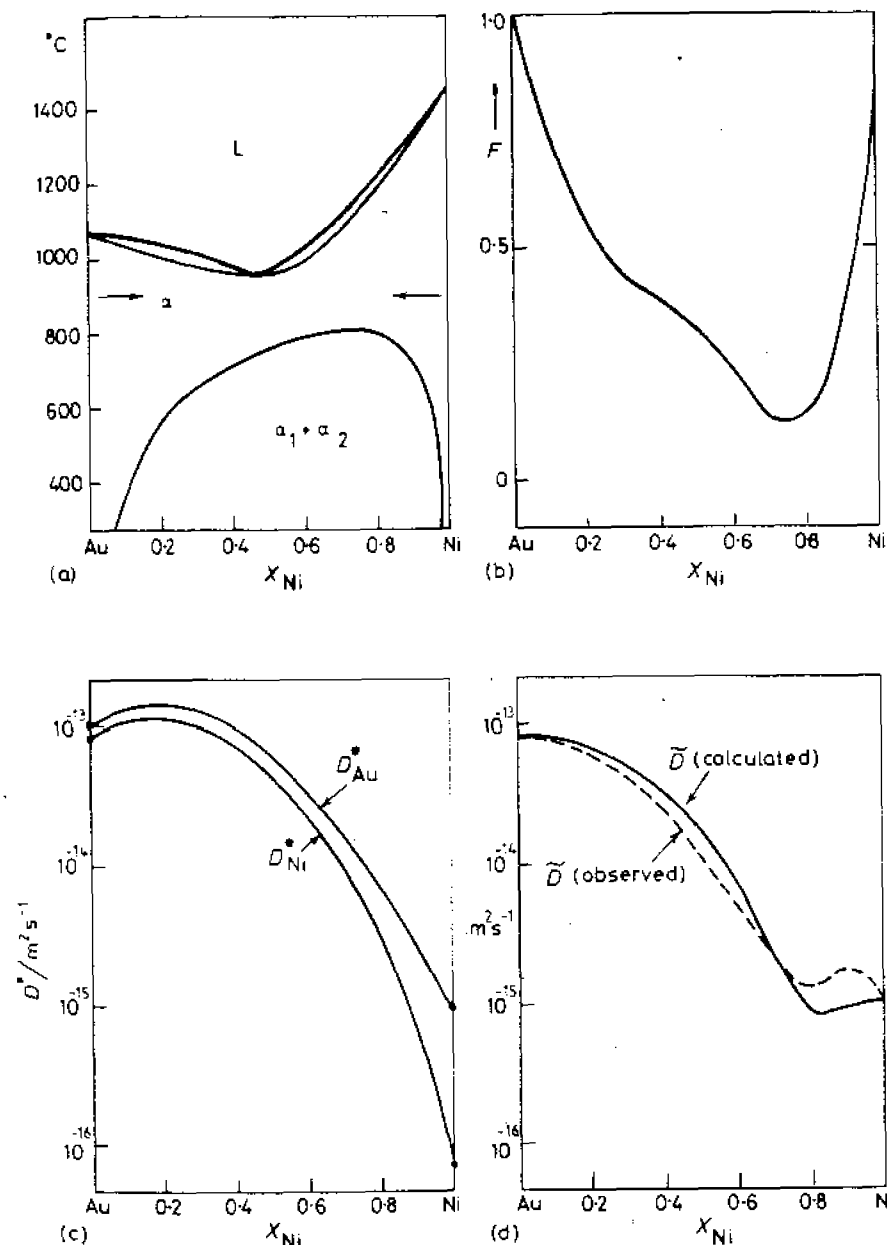


Fig. 2.22 Interdiffusion in Au-Ni alloys at 900 °C (a) Au-Ni phase diagram, (b) the thermodynamic factor, F , at 900 °C, (c) experimentally measured tracer diffusivities at 900 °C, (d) experimentally measured interdiffusion coefficients compared with values calculated from (b) and (c). (From J.E. Reynolds, B.L. Averbach and Morris Cohen, *Acta Metallurgica*, 5 (1957) 29.)

The second equality has been obtained by assuming M_B^* in the tracer experiment equals M_B in the chemical diffusion case. Substitution into Equations 2.64 and 2.51 therefore leads to the following relationships

$$\begin{aligned} D_A &= FD_A^* \\ D_B &= FD_B^* \end{aligned} \quad (2.70)$$

$$\text{and } \bar{D} = F(X_B D_A^* + X_A D_B^*) \quad (2.71)$$

where F is the thermodynamic factor, i.e.

$$F = \left\{ 1 + \frac{d \ln \gamma_A}{d \ln X_A} \right\} = \left\{ 1 + \frac{d \ln \gamma_B}{d \ln X_B} \right\} = \frac{X_A X_B}{RT} \frac{d^2 G}{dX^2} \quad (2.72)$$

The last equality follows from Equation 1.71.

In the case of the Au-Ni system, diffusion couple experiments have also been carried out so that data are available for the interdiffusion coefficient $\bar{D}$, the full line in Fig. 2.22d. In addition there is also enough thermodynamic data on this system for the thermodynamic factor F to be evaluated, Fig. 2.22b. It is therefore possible to check the assumption leading to the second equality in Equation 2.69 by combining the data in Fig. 2.22b and c using Equation 2.71. This produces the solid line in Fig. 2.22d. The agreement is within experimental error.

Before leaving Fig. 2.22 it is interesting to note how the diffusion coefficients are strongly composition dependent. There is a difference of about three orders of magnitude across the composition range. This can be explained by the lower liquidus temperature of the Au-rich compositions. Also in agreement with the rules of thumb given earlier, Au, with the lower melting temperature, diffuses faster than Ni at all compositions.

2.6 Diffusion in Ternary Alloys

The addition of a third diffusing species to a solid solution produces mathematical complexities which will not be considered here. Instead let us consider an illustrative example of some of the additional effects that can arise. Fe-Si-C alloys are particularly instructive for two reasons. Firstly silicon raises the chemical potential (or activity) of carbon in solution, i.e. carbon will not only diffuse from regions of high carbon concentration but also from regions rich in silicon. Secondly the mobilities of carbon and silicon are widely different. Carbon, being an interstitial solute, is able to diffuse far more rapidly than the substitutionally dissolved silicon.

Consider two pieces of steel, one containing 3.8% silicon and 0.48% carbon and the other 0.44% carbon but no silicon. If the two pieces are welded together and austenitized at 1050 °C, the carbon concentration profile shown in Fig. 2.23b is produced. The initial concentrations of silicon and carbon in the couple are shown in Fig. 2.23a and the resultant chemical potentials of carbon by the dotted line in Fig. 2.23c. Therefore carbon atoms

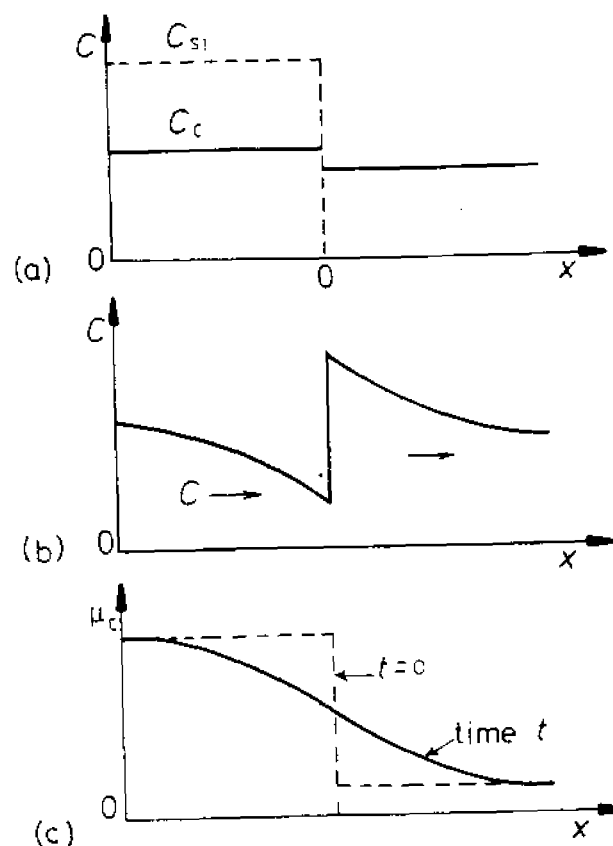


Fig. 2.23 (a) Carbon and silicon distribution in iron at $t = 0$. (b) Carbon distribution after high-temperature anneal. (c) Chemical potential of carbon v. distance.

on the silicon-rich side will jump over to the silicon-free side until the difference in concentration at the interface is sufficient to equalize the activity, or chemical potential, of carbon on both sides. The carbon atoms at the interface are therefore in *local equilibrium* and the interfacial compositions remain constant as long as the silicon atoms do not migrate. Within each half of the couple the silicon concentration gradients as shown in Fig. 2.23b. The resultant chemical potential varies smoothly across the whole specimen Fig. 2.23c. If the total length of the diffusion couple is sufficiently small the carbon concentration in each block will eventually equal the interfacial compositions and the chemical potential of carbon will be the same everywhere. The alloy is now in a state of partial equilibrium. It is only partial because the chemical potential of the silicon is not uniform. Given sufficient time the silicon atoms will also diffuse over significant distances and the carbon atoms will continually redistribute themselves to maintain a constant chemical potential. In the final equilibrium state the concentrations of carbon and

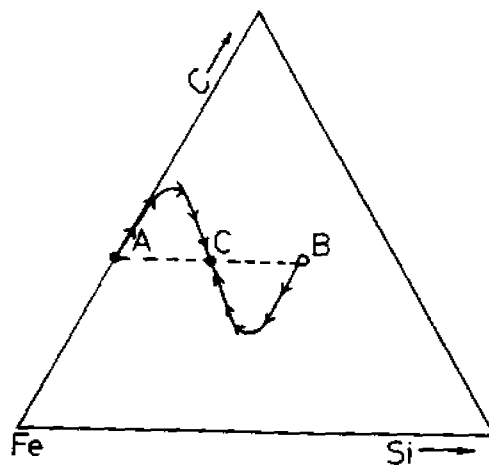


Fig. 2.24 Schematic diagram showing the change in composition of two points (A and B) on opposite sides of the diffusion couple in Fig. 2.23. C is the final equilibrium composition of the whole bar. (After L.S. Darken, *Trans. AIME*, 180 (1949) 430. © American Society for Metals and the Metallurgical Society of AIME, 1949.)

silicon are uniform everywhere. The change in composition of two points on opposite sides of the weld will be as illustrated on the ternary diagram of Fig. 2.24.

The redistribution of carbon in the Fe-Si-C system is particularly interesting since the mobilities of carbon and silicon are so different. Similar, though less striking effects can arise in ternary systems where all three components diffuse substitutionally if their diffusivities (or mobilities) are unequal.

2.7 High-Diffusivity Paths

In Section 2.4 the diffusion of atoms towards or away from dislocations, interfaces, grain boundaries and free surfaces was considered. In this section diffusion *along* these defects will be discussed. All of these defects are associated with a more open structure and it has been shown experimentally that the jump frequency for atoms migrating along these defects is higher than that for diffusion in the lattice. It will become apparent that under certain circumstances diffusion along these defects can be the dominant diffusion path.

2.7.1 Diffusion along Grain Boundaries and Free Surfaces

It is found experimentally that diffusion along grain boundaries and free surfaces can be described by

$$D_b = D_{b0} \exp \frac{-Q_b}{RT} \quad (2.73)$$

or

$$D_s = D_{s0} \exp \frac{-Q_s}{RT} \quad (2.74)$$

where D_b and D_s are the grain boundary and surface diffusivities and D_{b0} and D_{s0} are the frequency factors. Q_b and Q_s are the experimentally determined values of the activation energies for diffusion. In general, at any temperature the magnitudes of D_b and D_s relative to the diffusivity through defect-free lattice D_l are such that

$$D_s > D_b > D_l \quad (2.75)$$

This mainly reflects the relative ease with which atoms can migrate along free surfaces, interior boundaries and through the lattice. Surface diffusion can play an important role in many metallurgical phenomena, but in an average metallic specimen the total grain boundary area is much greater than the surface area so that grain boundary diffusion is usually most important.

The effect of grain boundary diffusion can be illustrated by considering a diffusion couple made by welding together two metals, A and B, as shown in Fig. 2.25. A atoms diffusing along the boundary will be able to penetrate much deeper than atoms which only diffuse through the lattice. In addition, as the concentration of solute builds up in the boundaries atoms will also diffuse

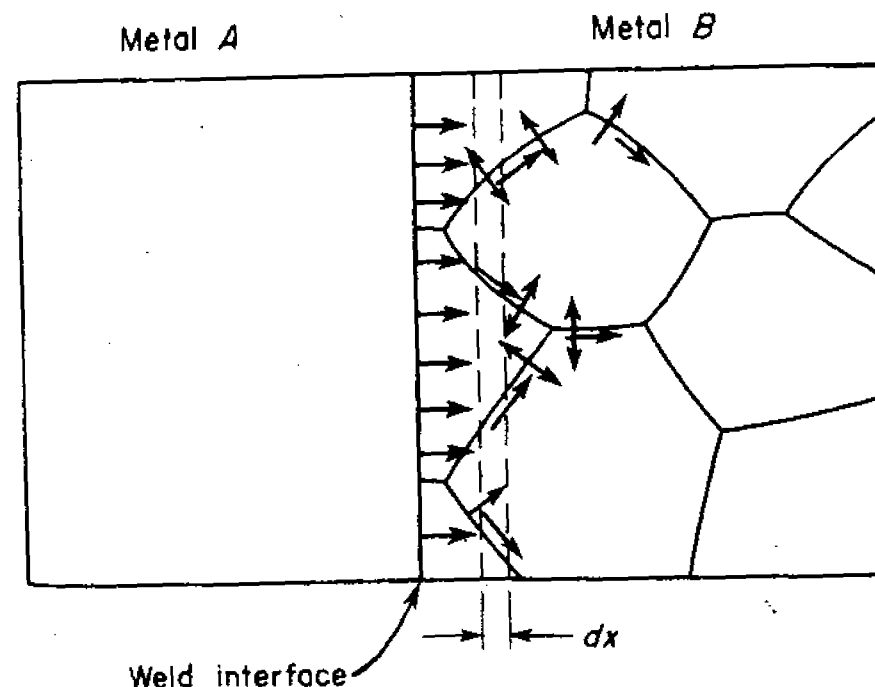


Fig. 2.25 The effect of grain boundary diffusion combined with volume diffusion. (After R.E. Reed-Hill, *Physical Metallurgy Principles*, 2nd edn., Van Nostrand, New York, 1973.)

from the boundary into the lattice. The process can be compared to the conduction of heat through a plastic in which a continuous network of aluminium sheets is embedded. The temperature at any point in such a specimen would be analogous to the concentration of solute in the diffusion couple. Points in the lattice close to grain boundaries can receive solute via the high conductivity path much more rapidly than if the boundaries were absent. Rapid diffusion along the grain boundaries increases the mean concentration in a slice such as dx in Fig. 2.25 and thereby produces an increase in the apparent diffusivity in the material as a whole. Consider now under what conditions grain boundary diffusion is important.

For simplicity let us take a case of steady-state diffusion through a sheet of material in which the grain boundaries are perpendicular to the sheet as shown in Fig. 2.26. Assuming that the concentration gradients in the lattice and along the boundary are identical, the fluxes of solute through the lattice J_l and along the boundary J_b will be given by

$$J_l = -D_l \frac{dC}{dx} \quad J_b = -D_b \frac{dC}{dx} \quad (2.76)$$

However the contribution of grain boundary diffusion to the total flux through the sheet will depend on the relative cross-sectional areas through which the solute is conducted.

If the grain boundary has an effective thickness δ and the grain size is d the total flux will be given by

$$J = (J_b \delta + J_l d) / d = - \left(\frac{D_b \delta + D_l d}{d} \right) \frac{dC}{dx} \quad (2.77)$$

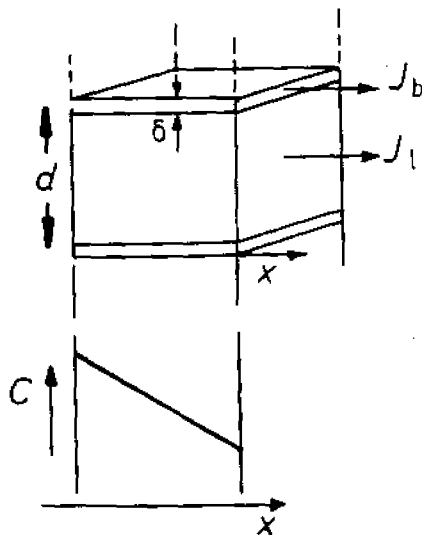


Fig. 2.26 Combined lattice and boundary fluxes during steady-state diffusion through a thin slab of material.

Thus the apparent diffusion coefficient in this case,

$$D_{app} = D_l + D_b \delta / d \quad (2.78)$$

$$\text{or} \quad \frac{D_{app}}{D_l} = 1 + \frac{D_b \delta}{D_l d} \quad (2.79)$$

It can be seen that the relative importance of lattice and grain boundary diffusion depends on the ratio $D_b \delta / D_l d$. When $D_b \delta \gg D_l d$ diffusion through the lattice can be ignored in comparison to grain boundary diffusion. Thus grain boundary diffusion makes a significant contribution to the total flux when

$$D_b \delta > D_l d \quad (2.80)$$

The effective width of a grain boundary is ~ 0.5 nm. Grain sizes on the other hand can vary from ~ 1 to $1000 \mu\text{m}$ and the effectiveness of the grain boundaries will vary accordingly. The relative magnitudes of $D_b \delta$ and $D_l d$ are most sensitive to temperature. This is illustrated in Fig. 2.27 which shows the effect of temperature on both D_l and D_b . Note that although $D_b > D_l$ at all temperatures the difference increases as temperature decreases. This is because the activation energy for diffusion along grain boundaries (Q_b) is lower than that for lattice diffusion (Q_l). For example, in fcc metals it is generally found that $Q_b \sim 0.5 Q_l$. This means that when the grain boundary diffusivity is scaled by the factor δ/d (Equation 2.78) the grain boundary contribution to the total, or apparent, diffusion coefficient is negligible in comparison to the lattice diffusivity at high temperatures, but dominates at low temperatures. In

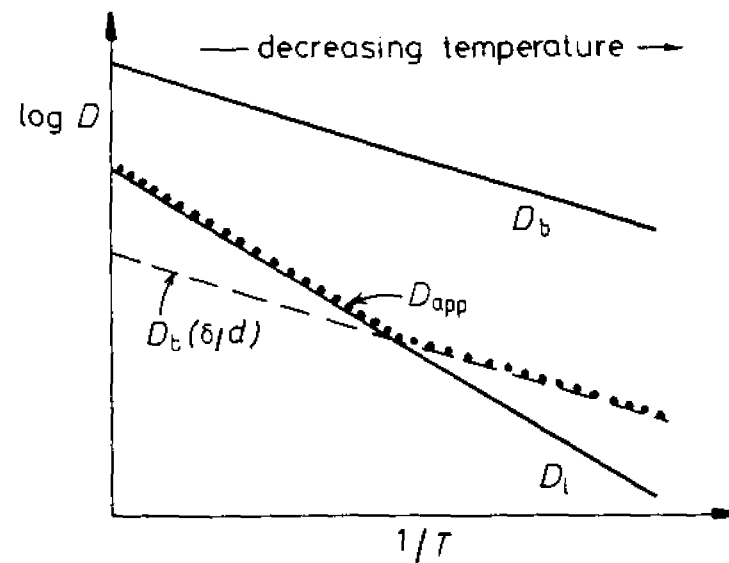


Fig. 2.27 Diffusion in a polycrystalline metal.

general it is found that grain boundary diffusion becomes important below about $0.75\text{--}0.8 T_m$, where T_m is the equilibrium melting temperature in degrees Kelvin.

The rate at which atoms diffuse along different boundaries is not the same, but depends on the atomic structure of the individual boundary. This in turn depends on the orientation of the adjoining crystals and the plane of the boundary. Also the diffusion coefficient can vary with direction within a given boundary plane. The reasons for these differences will become apparent in Chapter 3.

2.7.2 Diffusion along Dislocations

If grain boundary diffusion is compared to the conduction of heat through a material made of sheets of aluminium in a plastic matrix, the analogy for diffusion along dislocations would be aluminium wires in a plastic matrix. The dislocations effectively act as pipes along which atoms can diffuse with a diffusion coefficient D_p . The contribution of dislocations to the total diffusive flux through a metal will of course depend on the relative cross-sectional areas of pipe and matrix. Using the simple model illustrated in Fig. 2.28 it can easily be shown that the apparent diffusivity through a single crystal containing dislocations, D_{app} , is related to the lattice diffusion coefficient by

$$\frac{D_{app}}{D_l} = 1 + g \cdot \frac{D_p}{D_l} \quad (2.81)$$

where g is the cross-sectional area of 'pipe' per unit area of matrix. In a well-annealed material there are roughly 10^5 dislocations mm^{-2} . Assuming that the cross-section of a single pipe accommodates about 10 atoms while the matrix contains about 10^{13} atoms mm^{-2} , makes $g \approx 10^{-7}$.

At high temperatures diffusion through the lattice is rapid and gD_p/D_l is very small so that the dislocation contribution to the total flux of atoms is negligible. However, since the activation energy for pipe diffusion is less than

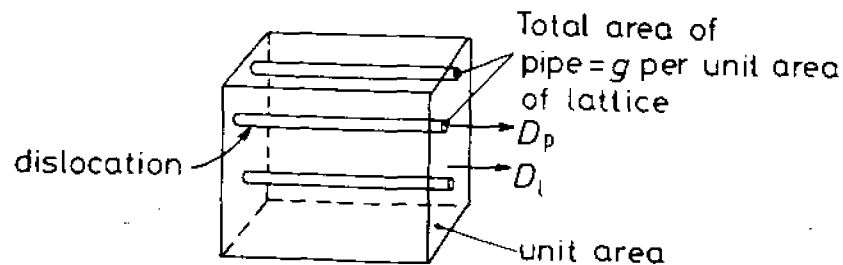


Fig. 2.28 Dislocations act as a high conductivity path through the lattice.

for lattice diffusion, D_l decreases much more rapidly than D_p with decreasing temperature, and at low temperatures gD_p/D_l can become so large that the apparent diffusivity is entirely due to diffusion along dislocations.

2.8 Diffusion in Multiphase Binary Systems

So far only diffusion in single-phase systems has been considered. In most practical cases, however, diffusion occurs in the presence of more than one phase. For example diffusion is involved in solidification transformations and diffusional transformations in solids (Chapters 4 and 5). Another example of multiphase diffusion arises when diffusion couples are made by welding together two metals that are not completely miscible in each other. This situation arises in practice with galvanized iron and hot-dipped tin plate for example. In order to understand what happens in these cases consider the hypothetical phase diagram in Fig. 2.29a. A diffusion couple made by welding together pure A and pure B will result in a layered structure containing α , β and γ . Annealing at temperature T_1 will produce a phase distribution and composition profile as shown in Fig. 2.29b. Usually X_B varies as shown from 0 to a in the α phase, from b to c in the β phase, and from d to 1 in the γ phase, where a , b , c and d are the solubility limits of the phases at T_1 . The compositions a and b are seen to be the equilibrium compositions of the α and β phases in the $\alpha + \beta$ field of the phase diagram. The α and β phases are therefore in *local equilibrium* across the α/β interface. Similarly β and γ are in local equilibrium across the β/γ interface. A sketch of the free energy-composition diagram for this system at T_1 will show that the chemical potentials (or activities) of A and B will vary continuously across the diffusion couple. Figure 2.29c shows how the activity of B varies across the couple (see problem 2.8). Clearly the equilibrium condition $a_B^a = a_B^b$ is satisfied at the α/β interface (point p in Fig. 2.29c). Similar considerations apply for A and for the β/γ interface.

The α/β and β/γ interfaces are not stationary but move as diffusion progresses. For example if the overall composition of the diffusion couple lies between b and c the final equilibrium state will be a single block of β .

A complete solution of the diffusion equations for this type of diffusion couple is complex. However an expression for the rate at which the boundaries move can be obtained as follows. Consider the planar α/β interface as shown in Fig. 2.30. If unit area of the interface moves a distance dx a volume $(dx \cdot 1)$ will be converted from α containing C_B^a B-atoms m^{-3} to β containing C_B^b B-atoms m^{-3} . This means that a total of

$$(C_B^b - C_B^a)dx$$

B atoms must accumulate at the α/β interface (the shaded area in Fig. 2.30). There is a flux of B towards the interface from the β phase equal to

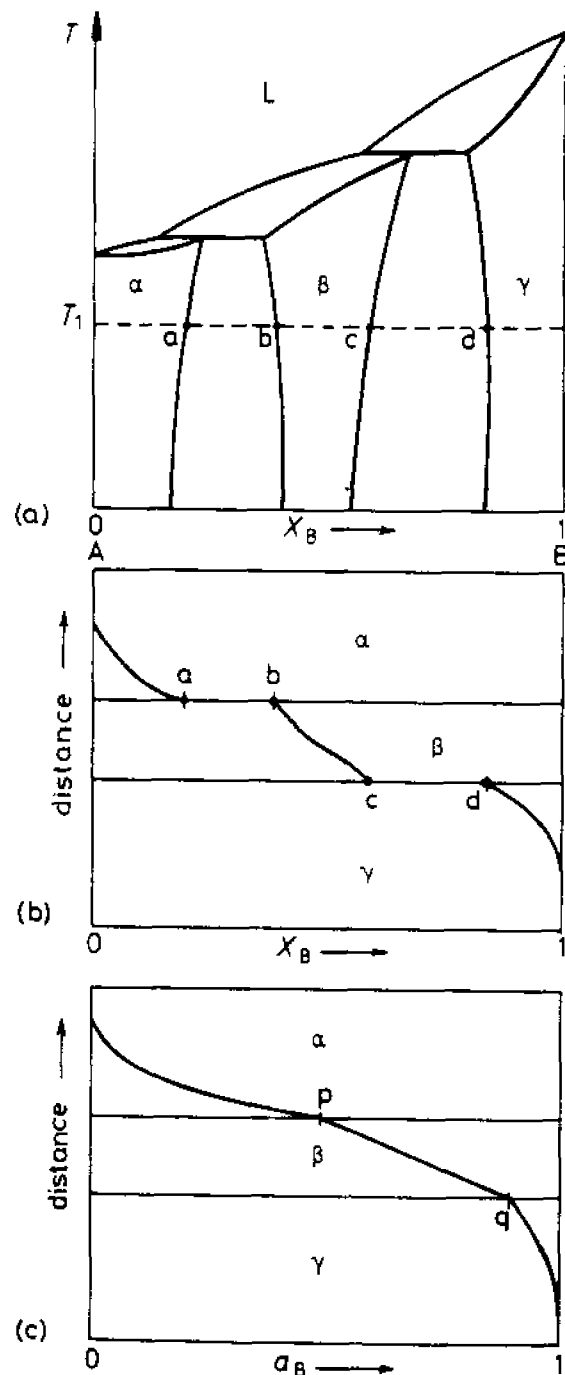


Fig. 2.29 (a) A hypothetical phase diagram. (b) A possible diffusion layer structure for pure A and B welded together and annealed at T_1 . (c) A possible variation of the activity of B (a_B) across the diffusion couple.

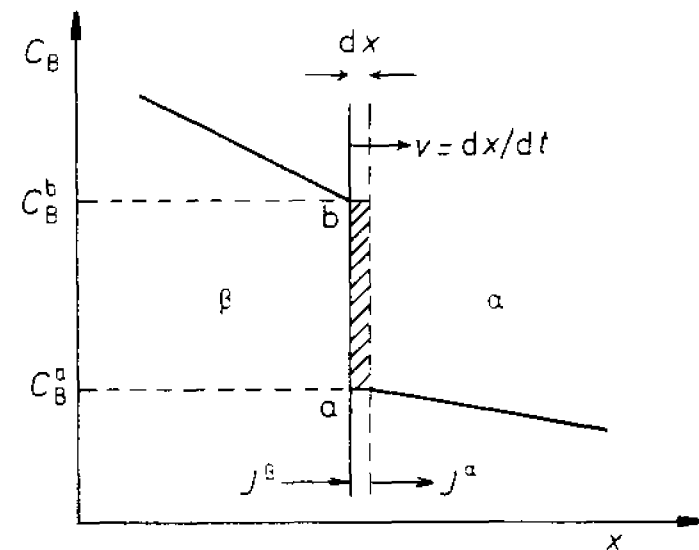


Fig. 2.30 Concentration profile across the α/β interface and its associated movement assuming diffusion control.

$-\bar{D}(\beta)\partial C_B^b/\partial x$ and a similar flux away from the interface into the α phase equal to $-\bar{D}(\alpha)\partial C_B^a/\partial x$. In a time dt , therefore, there will be an accumulation of B atoms given by

$$\left\{ \left(-\bar{D}(\beta) \frac{\partial C_B^b}{\partial x} \right) - \left(-\bar{D}(\alpha) \frac{\partial C_B^a}{\partial x} \right) \right\} dt$$

Equating the above expressions gives the instantaneous velocity of the α/β interface v as

$$v = \frac{dx}{dt} = \frac{1}{(C_B^b - C_B^a)} \left\{ \bar{D}(\alpha) \frac{\partial C_B^a}{\partial x} - \bar{D}(\beta) \frac{\partial C_B^b}{\partial x} \right\} \quad (2.82)$$

In the above treatment it has been assumed that the α/β interface moves as fast as allowed by the diffusive fluxes in the two adjacent phases. This is quite correct when the two phases are in local equilibrium, and is usually true in diffusion-couple experiments. However, it is not true for all moving inter-phase interfaces. By assuming local equilibrium at the interface it has also been assumed that atoms can be transferred across the interface as easily as they can diffuse through the matrix. Under these circumstances μ_B and a_B are continuous across the interface. However, in general this need not be true. If, for some reason, the interface has a low mobility the concentration difference across the boundary ($C_B^b - C_B^a$) will increase, thereby creating a discontinuity of chemical potential across the boundary. The problem of evaluating the boundary velocity in this case is more complex. Not only must the flux of atoms to the interface balance the rate of accumulation due to the boundary

migration and the rate of diffusion away into the other phase, but it must also balance with the rate of transfer across the interface. In extreme cases the *interface reaction*, as it is sometimes called, can be so slow that there are virtually no concentration gradients in the two phases. Under these circumstances the interface migration is said to be *interface controlled*. The subject of interface migration is treated further in Section 3.5.

References

1. See for example P.G. Shewmon, *Diffusion in Solids*, McGraw-Hill, New York, 1963, p. 47.
2. C. Wert, *Phys. Rev.*, **79**:601 (1950).
3. C. Wert, *J. Appl. Phys.*, **21**:1196 (1950).
4. E. Johnson and M. Hill, *Trans. AIME*, **218**:1104 (1960).
5. L.S. Darken, *Trans. Met. Soc. AIME*, **175**:184 (1948).
6. For details of these methods see J. Crank, *The Mathematics of Diffusion*, Oxford University Press, 1956.
7. A.D. Smigelskas and E.O. Kirkendall, *Trans. Met. Soc. AIME*, **171**:130 (1947).
8. P.G. Shewmon, *Diffusion in Solids*, McGraw-Hill, New York, 1963, p. 134.
9. See for example A.G. Guy *Introduction to Materials Science*, McGraw-Hill, 1971, p. 284.
10. J.E. Reynolds, B.L. Averbach and M. Cohen, *Acta Metallurgica*, **5**:29 (1957).

Further Reading

- A.M. Brown and M.F. Ashby, 'Correlations for Diffusion Constants', *Acta Metallurgica*, **28**:1085 (1980).
- C.P. Flynn, *Point Defects and Diffusion*, Oxford University Press, 1972.
- S. Mrowec, *Defects and Diffusion in Solids—An Introduction*, Elsevier, Amsterdam, 1980.
- P.G. Shewmon, *Diffusion in Solids*, 2nd edition, McGraw-Hill, New York, 1989.
- J.L. Bocquet, G. Brébec and Y. Limoge, 'Diffusion in metals and alloys', Chapter 8 in *Physical Metallurgy*, R.W. Cahn and P. Haasen (Eds.), North-Holland, 1983.

Exercises

- 2.1 A thin sheet of iron is in contact with a carburizing gas on one side and a decarburizing gas on the other at temperature of 1000 °C.

- (a) Sketch the resultant carbon concentration profile when a steady state has been reached assuming the surface concentrations are maintained at 0.15 and 1.4 wt% C.
- (b) If D_c increases from $2.5 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ at 0.15% C to $7.7 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ at 1.4% C what will be the quantitative relationship between the concentration gradients at the surfaces?
- (c) Estimate an approximate value for the flux of carbon through the sheet if the thickness is 2 mm (0.8 wt% C = 60 kg m^{-3} at 1000 °C).
- 2.2 It was stated in Section 2.2.1 that $D = \Gamma\alpha^2/6$ applies to any diffusing species in any cubic lattice. Show that this is true for vacancy diffusion in a pure fcc metal. (Hint: consider two adjacent {111} planes and determine what fraction of all possible jumps result in the transfer of a vacancy between the two planes. Is the same result obtained by considering adjacent {100} planes?)
- 2.3 A small quantity of radioactive gold was deposited on the end of a gold cylinder. After holding for 24 h at a high temperature the specimen was sectioned and the radioactivity of each slice was as follows:

Distance from end of bar

to centre of slice/ μm :	10	20	30	40	50
Activity:	83.8	66.4	42.0	23.6	8.74

Use the data to determine D .

- 2.4 Prove by differentiation that Equation 2.20 is a solution of Fick's second law.
- 2.5 Fourier analysis is a powerful tool for the solution of diffusion problems when the initial concentration profile is not sinusoidal. Consider for example the diffusion of hydrogen from an initially uniform sheet of iron. If the concentration outside the sheet is maintained at zero the resultant concentration profile is initially a top-hat function. Fourier analysis of this function shows that it can be considered as an infinite series of sine terms:

$$C(x) = \frac{4C_0}{\pi} \sum_{i=0}^{\infty} \frac{1}{2i+1} \sin \frac{(2i+1)\pi x}{l}$$

where l is the thickness of the sheet and C_0 is the initial concentration.

- (a) Plot the first two terms of this series. If during diffusion the surface concentration is maintained close to zero each Fourier component can be considered to decrease exponentially with time with a time constant $\tau_i = l^2/(2i+1)^2\pi^2 D$. The solution to the diffusion equation therefore becomes

$$C(x, t) = \frac{4C_0}{\pi} \sum_{i=0}^{\infty} \frac{1}{2i+1} \sin \left\{ \frac{(2i+1)\pi x}{l} \right\} \exp \left(-\frac{t}{\tau_i} \right)$$

- (b) Derive an equation for the time at which the amplitude of the second term is less than 5% of the first term.
- (c) Approximately how long will it take to remove 95% of all the hydrogen from an initially uniform plate of α -iron at 20 °C if (i) the plate is 10 mm thick and if (ii) it is 100 mm thick, assuming the surface concentration is maintained constant at zero? (Use data in Table 2.1.)

2.6 Figure 2.31 shows the molar free energy–composition diagram for the A–B system at temperature T_1 . Imagine that a block of α with composition (1) is welded to a block of β phase with composition (2). By considering the chemical potentials of the A and B atoms in both the α and β phases predict which way the atoms will move during a diffusion anneal at T_1 . Show that this leads to a reduction of the molar free energy of the couple. Indicate the compositions of the two phases when equilibrium is reached.

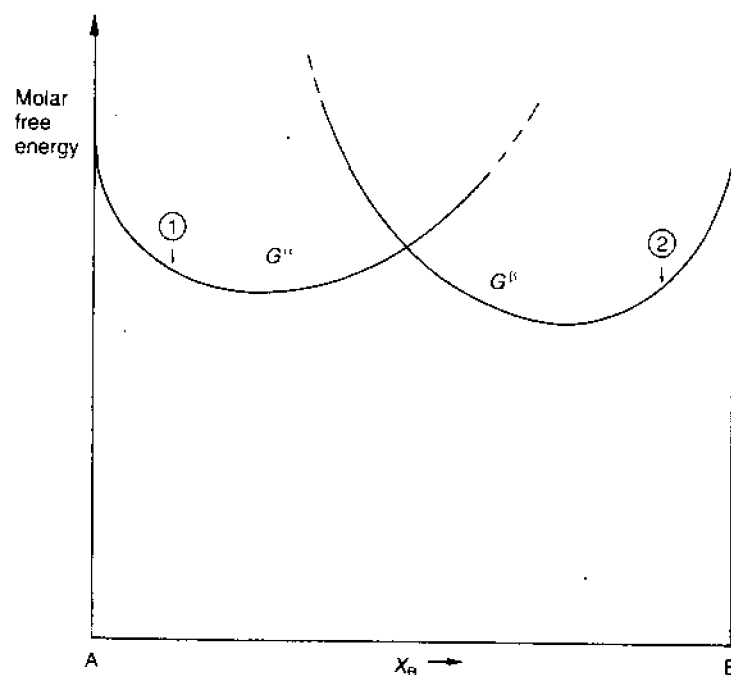


Fig. 2.31

2.7 A diffusion couple including inert wires was made by plating pure copper on to a block of α -brass with a composition Cu–30 wt% Zn, Fig. 2.20. After 56 days at 785°C the marker velocity was determined as $2.6 \times 10^{-8} \text{ mm s}^{-1}$. Microanalysis showed that the composition at the markers was $X_{\text{Zn}} = 0.22$, $X_{\text{Cu}} = 0.78$, and that $\partial X_{\text{Zn}}/\partial x$ was 0.089 mm^{-1} .

From an analysis of the complete penetration curve $\bar{D}^a$ at the markers was calculated as $4.5 \times 10^{-13} \text{ m}^2 \text{ s}^{-1}$. Use this data to calculate D_{Zn}^a and D_{Cu}^a in brass at 22 atomic % Zn. How would you expect D_{Zn}^a , D_{Cu}^a and $\bar{D}^a$ to vary as a function of composition?

- 2.8 Draw possible free energy–composition curves for the system in Fig. 2.29 at T_1 . Derive from this a μ_B – X_B and an a_B – X_B diagram (similar to Fig. 1.28). Mark the points corresponding to p and q in Fig. 2.29c. Sketch diagrams similar to Fig. 2.29c to show a_A , μ_A and μ_B across the diffusion couple. What will be the final composition profile when the couple reaches equilibrium if the overall composition lies (i) between a and b, (ii) below a?
- 2.9 Figure 2.32 is a hypothetical phase diagram for the A–B system. At a temperature T_1 B is practically insoluble in A, whereas B can dissolve 10 atomic % A. A diffusion couple made by welding together pure A and pure B is annealed at T_1 . Show by a series of sketches how the concentration profiles and α/β interface position will vary with time. If the overall composition of the couple is 50 atomic % B what will be the maximum displacement of the α/β interface? (Assume α and β have equal molar volumes.)

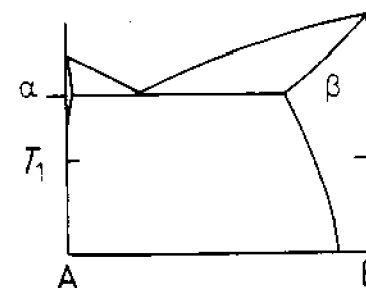


Fig. 2.32

3

Crystal Interfaces and Microstructure

Basically three different types of interface are important in metallic systems:

1. The free surfaces of a crystal (solid/vapour interface)
2. Grain boundaries (α/α interfaces)
3. Interphase interfaces (α/β interfaces).

All crystals possess the first type of interface. The second type separates crystals with essentially the same composition and crystal structure, but a different orientation in space. The third interface separates two different phases that can have different crystal structures and/or compositions and therefore also includes solid/liquid interfaces.

The great majority of phase transformations in metals occur by the growth of a new phase (β) from a few nucleation sites within the parent phase (α)—a nucleation and growth process. The α/β interface therefore plays an important role in determining the kinetics of phase transformations and is the most important class of interface listed. It is, however, also the most complex and least understood, and this chapter thus begins by first considering the simpler interfaces, (1) and (2).

The solid/vapour interface is of course itself important in vaporization and condensation transformations, while grain boundaries are important in recrystallization, i.e. the transformation of a highly deformed grain structure into new undeformed grains. Although no new phase is involved in recrystallization it does have many features in common with phase transformations.

The importance of interfaces is not restricted to what can be called the primary transformation. Since interfaces are an almost essential feature of the transformed microstructure, a second (slower) stage of most transformations is the microstructural coarsening that occurs with time¹. This is precisely analogous to the grain coarsening or grain growth that follows a recrystallization transformation.

3.1 Interfacial Free Energy

It is common practice to talk of interfacial energy. In reality, however, what is usually meant and measured by experiment is the interfacial free energy, γ . The free energy of a system containing an interface of area A and free energy

γ per unit area is given by

$$G = G_0 + A\gamma \quad (3.1)$$

where G_0 is the free energy of the system assuming that all material in the system has the properties of the bulk— γ is therefore the excess free energy arising from the fact that some material lies in or close to the interface. It is also the work that must be done at constant T and P to create unit area of interface.

Consider for simplicity a wire frame suspending a liquid film, Fig. 3.1. If one bar of the frame is movable it is found that a force F per unit length must be applied to maintain the bar in position. If this force moves a small distance so that the total area of the film is increased by dA the work done by the force is FdA . This work is used to increase the free energy of the system by dG . From Equation 3.1

$$dG = \gamma dA + A d\gamma$$

Equating this with FdA gives

$$F = \gamma + A \frac{d\gamma}{dA} \quad (3.2)$$

In the case of a liquid film the surface energy is independent of the area of the interface and $d\gamma/dA = 0$. This leads to the well-known result

$$F = \gamma \quad (3.3)$$

i.e. a surface with a free energy $\gamma \text{ J m}^{-2}$ exerts a surface tension of $\gamma \text{ N m}^{-1}$.

In the case of interfaces involving *solids*, however, it is not immediately obvious that γ is independent of area. Since a liquid is unable to support shear stresses, the atoms within the liquid can rearrange during the stretching process and thereby maintain a constant surface structure. Solids, however, are much more viscous and the transfer of atoms from the bulk to the surface, which is necessary to maintain an unchanged surface structure and energy, will take much longer. If this time is long in comparison to the time of the experiment then $d\gamma/dA \neq 0$ and surface free energy and surface tension will not be identical. Nevertheless, at temperatures near the melting point the atomic mobility is usually high enough for Equation 3.3 to be applicable.

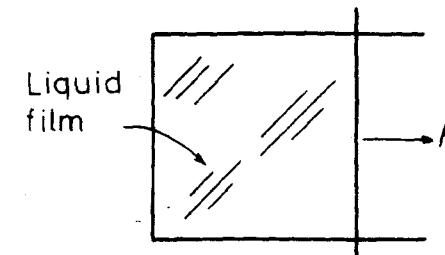


Fig. 3.1 A liquid film on a wire frame.

3.2 Solid/Vapour Interfaces

To a first approximation the structure of solid surfaces can be discussed in terms of a hard sphere model. If the surface is parallel to a low-index crystal plane the atomic arrangement will be the same as in the bulk, apart from perhaps a small change in lattice parameter. (This assumes that the surface is uncontaminated: in real systems surfaces will reduce their free energies by the adsorption of impurities.) Figure 3.2 for example shows the {111} {200} {220} atom planes in the fcc metals. Note how the density of atoms in these planes decreases as $(h^2 + k^2 + l^2)$ increases. (The notation {200} and {220} has been used instead of {100} and {110} because the spacing of equivalent atom planes is then given by $a/\sqrt{(h^2 + k^2 + l^2)}$ where a is the lattice parameter.)

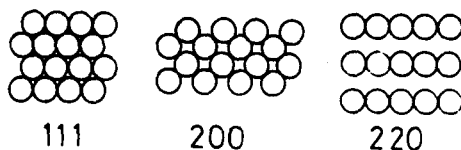


Fig. 3.2 Atomic configurations on the three closest-packed planes in fcc crystals: (111), (200) and (220).

The origin of the surface free energy is that atoms in the layers nearest the surface are without some of their neighbours. Considering only nearest neighbours it can be seen that the atoms on a {111} surface, for example, are deprived of three of their twelve neighbours. If the bond strength of the metal is ϵ each bond can be considered as lowering the internal energy of each atom by $\epsilon/2$. Therefore every surface atom with three 'broken bonds' has an excess internal energy of $3\epsilon/2$ over that of the atoms in the bulk. For a pure metal ϵ can be estimated from the heat of sublimation L_s . (The latent heat of sublimation is equal to the sum of the latent heat of melting (or fusion) and the latent heat of vaporization.) If 1 mol of solid is vaporized $12 N_a$ broken bonds are formed. Therefore $L_s = 12 N_a \epsilon/2$. Consequently the energy of a {111} surface should be given by

$$E_{sv} = 0.25 L_s / N_a \quad \text{J/surface atom} \quad (3.4)$$

This result will only be approximate since second nearest neighbours have been ignored and it has also been assumed that the strengths of the remaining bonds in the surface are unchanged from the bulk values.

From the definition of Gibbs free energy the surface free energy will be given by

$$\gamma = E + PV - TS \quad (3.5)$$

Thus even if the 'PV' term is ignored surface entropy effects must be taken into account. It might be expected that the surface atoms will have more freedom of movement and therefore a higher thermal entropy compared to

atoms in the bulk. Extra configurational entropy can also be introduced into the surface by the formation of surface vacancies for example. The surface of a crystal should therefore be associated with a positive excess entropy which will partly compensate for the high internal energy of Equation 3.4.

Experimental determination of γ_{sv} is difficult² but the measured values for pure metals indicate that near the melting temperature the surface free energy averaged over many surface planes is given by

$$\gamma_{sv} = 0.15 L_s / N_a \quad \text{J/surface atom} \quad (3.6)$$

As a result of entropy effects γ_{sv} is slightly dependent on temperature. From Equation 1.10

$$\left(\frac{\partial \gamma}{\partial T} \right)_P = -S \quad (3.7)$$

Measured values of S are positive and vary between 0 and $3 \text{ mJ m}^{-2} \text{ K}^{-1}$. Some selected values of γ_{sv} at the melting point are listed in Table 3.1. Note that metals with high melting temperatures have high values for L_s and high surface energies.

Table 3.1 Average Surface Free Energies of Selected Metals

Values selected from H. Jones 'The surface energy of solid metals', *Metal Science Journal*, 5:15 (1971). Experimental errors are generally about 10%. The values have been extrapolated to the melting temperature, T_m .

Crystal	$T_m / ^\circ\text{C}$	$\gamma_{sv} / \text{mJ m}^{-2}$
Sn	232	680
Al	660	1080
Ag	961	1120
Au	1063	1390
Cu	1084	1720
δ -Fe	1536	2080
Pt	1769	2280
W	3407	2650

It can be seen from the above simple model that different crystal surfaces should have different values for E_{sv} depending on the number of broken bonds (see exercise 3.1). A little consideration will show that for the surfaces shown in Fig. 3.2 the number of broken bonds at the surface will increase through the series {111} {200} {220}. Therefore ignoring possible differences in the entropy terms γ_{sv} should also increase along the same series.

When the macroscopic surface plane has a high or irrational $\{hkl\}$ index the surface will appear as a stepped layer structure where each layer is a close-packed plane. This is illustrated for a simple cubic crystal in Fig. 3.3.

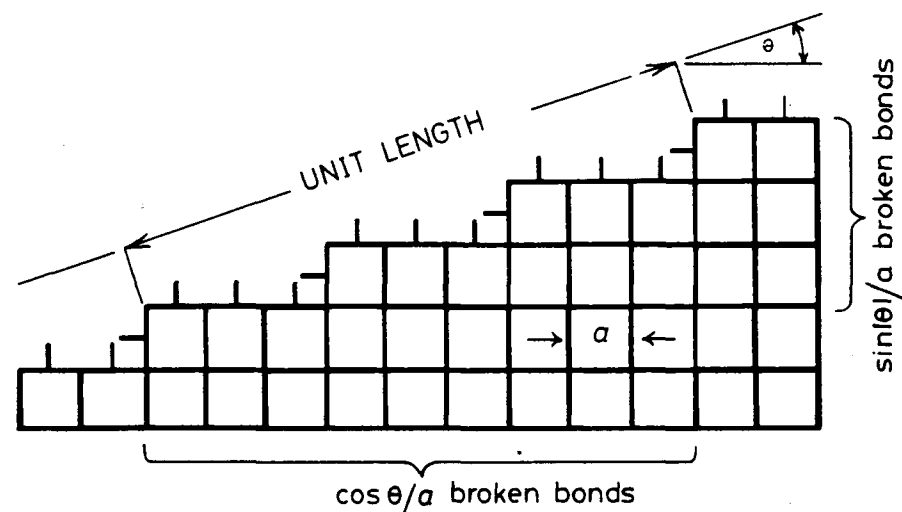


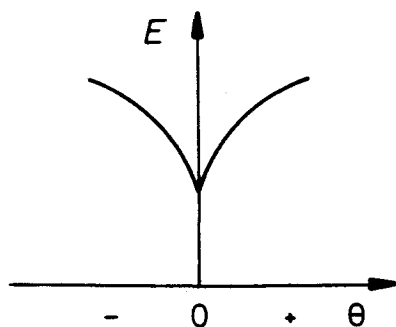
Fig. 3.3 The 'broken-bond' model for surface energy.

A crystal plane at an angle θ to the close-packed plane will contain broken bonds in excess of the close-packed plane due to the atoms at the steps. For unit length of interface in the plane of the diagram and unit length out of the paper (parallel to the steps) there will be $(\cos \theta/a)(1/a)$ broken bonds out of the close-packed plane and $(\sin |\theta|/a)(1/a)$ additional broken bonds from the atoms on the steps. Again attributing $\epsilon/2$ energy to each broken bond, then

$$E_{sv} = (\cos \theta + \sin |\theta|)\epsilon/2a^2 \quad (3.8)$$

This is plotted as a function of θ in Fig. 3.4. Note that the close-packed orientation ($\theta = 0$) lies at a cusped minimum in the energy plot. Similar arguments can be applied to any crystal structure for rotations about any axis from any reasonably close-packed plane. All low-index planes should therefore be located at low-energy cusps.

If γ is plotted v. θ similar cusps are found, but as a result of entropy effects

Fig. 3.4 Variation of surface energy as a function of θ in Fig. 3.3.

they are less prominent than in the E - θ plot, and for the higher index planes they can even disappear.

A convenient method for plotting the variation of γ with surface orientation in three dimensions is to construct a surface about an origin such that the free energy of any plane is equal to the distance between the surface and the origin when measured along the normal to the plane in question. A section through such a surface is shown in Fig. 3.5a. This type of polar representation of γ is known as a γ -plot and has the useful property of being able to predict the equilibrium shape of an isolated single crystal.

For an isolated crystal bounded by several planes A_1, A_2 , etc. with energies γ_1, γ_2 , etc. the total surface energy will be given by the sum $A_1\gamma_1 + A_2\gamma_2 + \dots$. The equilibrium shape has the property that $\sum A_i\gamma_i$ is a minimum and the shape that satisfies this condition is given by the following, so-called *Wulff construction*³. For every point on the γ surface, such as A in Fig. 3.5a, a plane is drawn through the point and normal to the radius vector

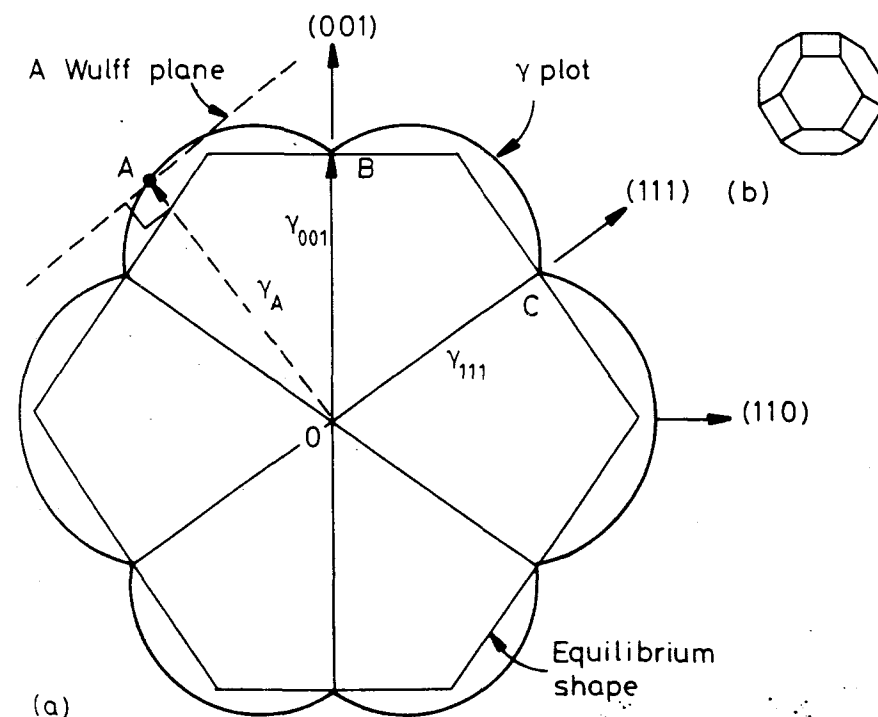


Fig. 3.5 (a) A possible $(1\bar{1}0)$ section through the γ -plot of an fcc crystal. The length OA represents the free energy of a surface plane whose normal lies in the direction OA. Thus $OB = \gamma_{(001)}$, $OC = \gamma_{(111)}$, etc. Wulff planes are those such as that which lies normal to the vector OA. In this case the Wulff planes at the cusps (B, C, etc.) give the inner envelope of all Wulff planes and thus the equilibrium shape. (b) The equilibrium shape in three dimensions showing $\{100\}$ (square faces) and $\{111\}$ (hexagonal faces).

OA. The equilibrium shape is then simply the inner envelope of all such planes. Therefore when the γ -plot contains sharp cusps the equilibrium shape is a polyhedron with the largest facets having the lowest interfacial free energy.

Equilibrium shapes can be determined experimentally by annealing small single crystals at high temperatures in an inert atmosphere, or by annealing small voids inside a crystal⁴. Fcc crystals for example usually assume a form showing $\{100\}$ and $\{111\}$ facets as shown in Fig. 3.5b. Of course when γ is isotropic, as for liquid droplets, both the γ -plot and equilibrium shapes are spheres.

When the equilibrium shape is known it is possible to use the Wulff theorem in reverse to give the relative interfacial free energies of the observed facet planes. In Fig. 3.5 for example the widths of the crystal in the $\langle 111 \rangle$ and $\langle 100 \rangle$ directions will be in the ratio of $\gamma(111) : \gamma(100)$. $\{110\}$ facets are usually missing from the equilibrium shape of fcc metals, but do however appear for bcc metals⁵.

The aim of this section has been to show, using the simplest type of interface, the origin of interfacial free energy, and to show some of the methods available for estimating this energy. Let us now consider the second type of interface, grain boundaries.

3.3 Boundaries in Single-Phase Solids

The grains in a single-phase polycrystalline specimen are generally in many different orientations and many different types of grain boundary are therefore possible. The nature of any given boundary depends on the misorientation of the two adjoining grains and the orientation of the boundary plane relative to them. The lattices of any two grains can be made to coincide by rotating one of them through a suitable angle about a *single* axis. In general the axis of rotation will not be simply oriented with respect to either grain or the grain-boundary plane, but there are two special types of boundary that are relatively simple. These are pure tilt boundaries and pure twist boundaries, as illustrated in Fig. 3.6. A tilt boundary occurs when the axis of rotation is parallel to the plane of the boundary, Fig. 3.6a, whereas a twist boundary is formed when the rotation axis is perpendicular to the boundary, Fig. 3.6b.

3.3.1 Low-Angle and High-Angle Boundaries

It is simplest to first consider what happens when the misorientation between two grains is small. This type of boundary can be simply considered as an array of dislocations. Two idealized boundaries are illustrated in Fig. 3.7. These are symmetrical *low-angle tilt* and *low-angle twist* boundaries. The

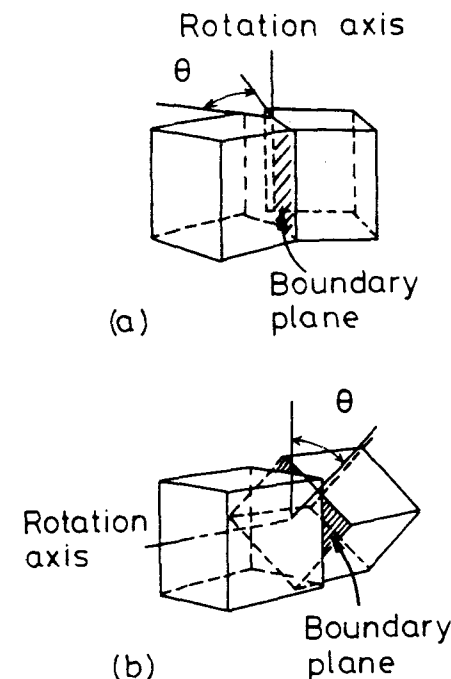


Fig. 3.6 The relative orientations of the crystals and the boundary forming (a) a tilt boundary (b) a twist boundary.

low-angle tilt boundary is an array of parallel edge dislocations, whereas the twist boundary is a cross-grid of two sets of screw dislocations. In each case the atoms in the regions between the dislocations fit almost perfectly into both adjoining crystals whereas the dislocation cores are regions of poor fit in which the crystal structure is highly distorted.

The tilt boundary need not be symmetrical with respect to the two adjoining crystals. However, if the boundary is unsymmetrical dislocations with different Burgers vectors are required to accommodate the misfit, as illustrated in Fig. 3.8. In general boundaries can be a mixture of the tilt and twist type in which case they must contain several sets of different edge and screw dislocations.

The energy of a low-angle grain boundary is simply the total energy of the dislocations within unit area of boundary. (For brevity the distinction between internal energy and free energy will usually not be made from now on except where essential to understanding.) This depends on the spacing of the dislocations which, for the simple arrays in Fig. 3.7, is given by

$$D = \frac{b}{\sin \theta} \approx \frac{b}{\theta} \quad (3.9)$$

where b is the Burgers vector of the dislocations and θ is the angular misorientation across the boundary. At very small values of θ the dislocation

spacing is very large and the grain boundary energy γ is approximately proportional to the density of dislocations in the boundary ($1/D$), i.e.

$$\gamma \propto \theta \quad (3.10)$$

However as θ increases the strain fields of the dislocations progressively cancel out so that γ increases at a decreasing rate as shown in Fig. 3.9. In general when θ exceeds $10\text{--}15^\circ$ the dislocation spacing is so small that the dislocation cores overlap and it is then impossible to physically identify the individual dislocations (see Fig. 3.10). At this stage the grain-boundary energy is almost independent of misorientation, Fig. 3.9.

When $\theta > 10\text{--}15^\circ$ the boundary is known as a *random high-angle grain boundary*. The difference in structure between low-angle and high-angle grain boundaries is lucidly illustrated by the bubble-raft model in Fig. 3.11. High-angle boundaries contain large areas of poor fit and have a relatively open structure. The bonds between the atoms are broken or highly distorted and consequently the boundary is associated with a relatively high energy. In

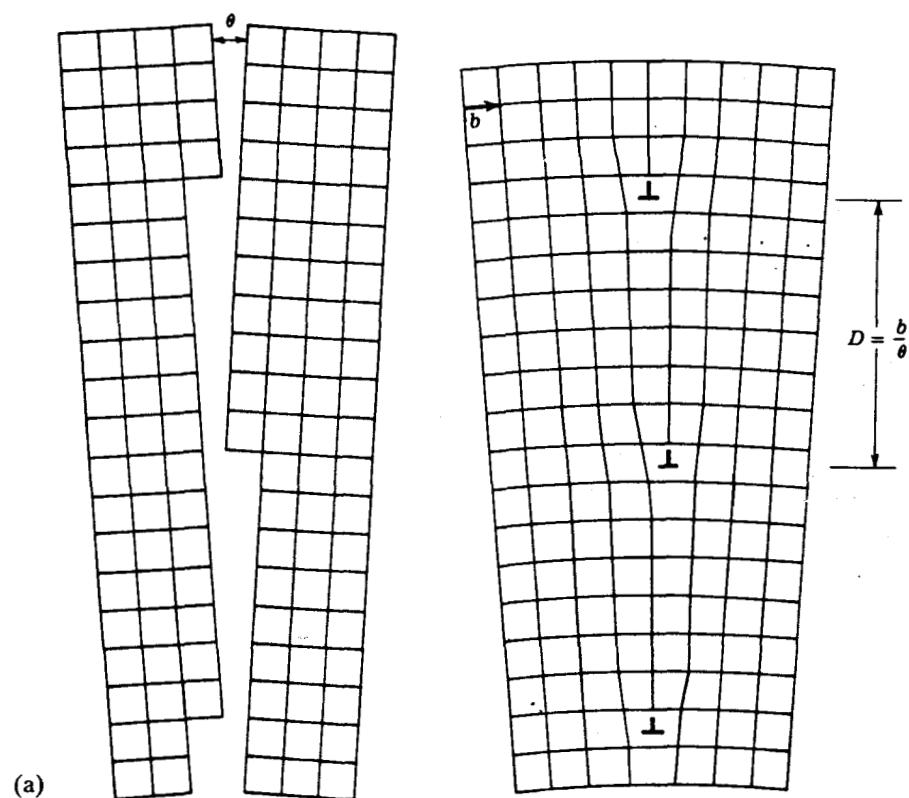


Fig. 3.7 (a) Low-angle tilt boundary, (b) low-angle twist boundary: $\circ$ atoms in crystal below boundary, $\bullet$ atoms in crystal above boundary. (After W.T. Read Jr., *Dislocations in Crystals*, McGraw-Hill, New York, 1953.)

low-angle boundaries, however, most of the atoms fit very well into both lattices so that there is very little free volume and the interatomic bonds are only slightly distorted. The regions of poor fit are restricted to the dislocation cores which are associated with a higher energy similar to that of the random high-angle boundary.

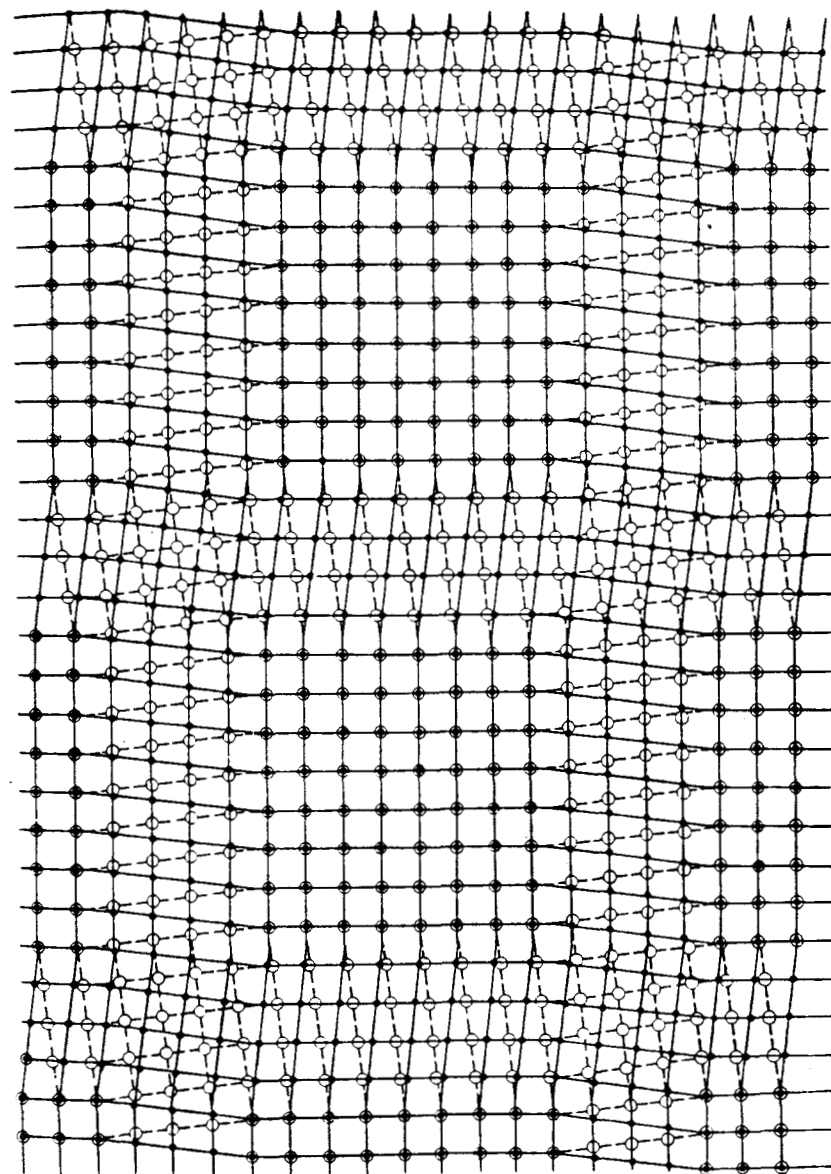


Fig. 3.7 (b)

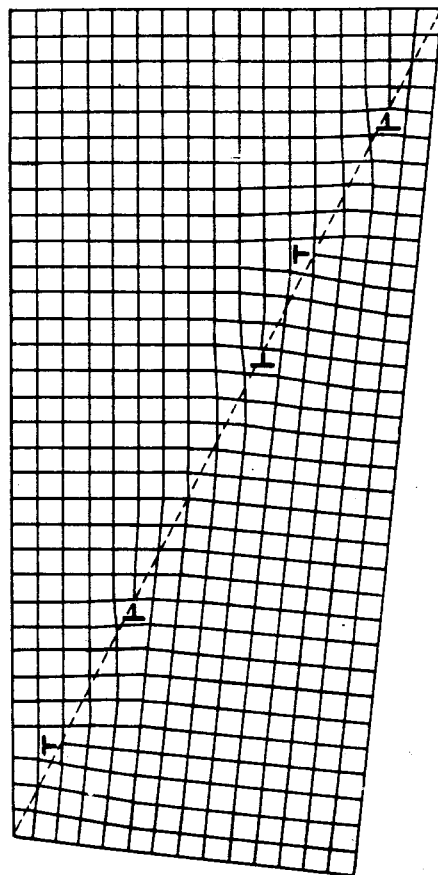


Fig. 3.8 An unsymmetric tilt boundary. Dislocations with two different Burgers vectors are present. (After W.T. Read Jr., *Dislocations in Crystals*, McGraw-Hill, New York, 1953.)

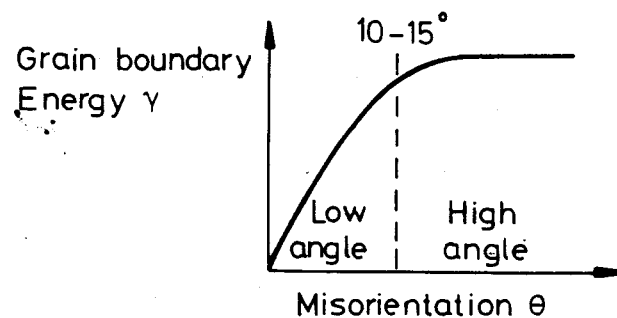


Fig. 3.9 Variation of grain boundary energy with misorientation (schematic).

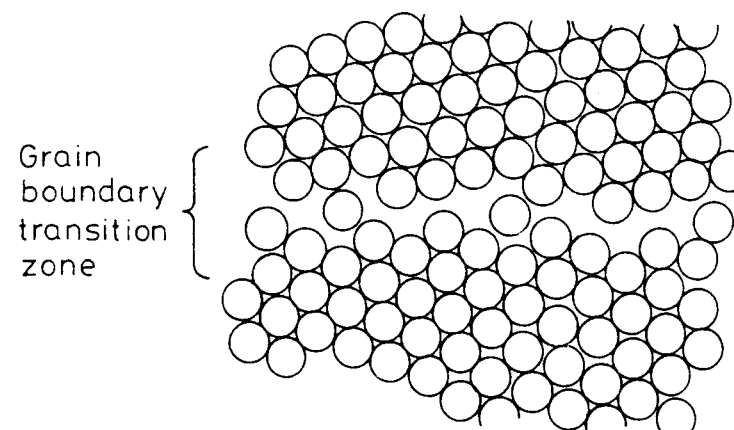


Fig. 3.10 Disordered grain boundary structure (schematic).

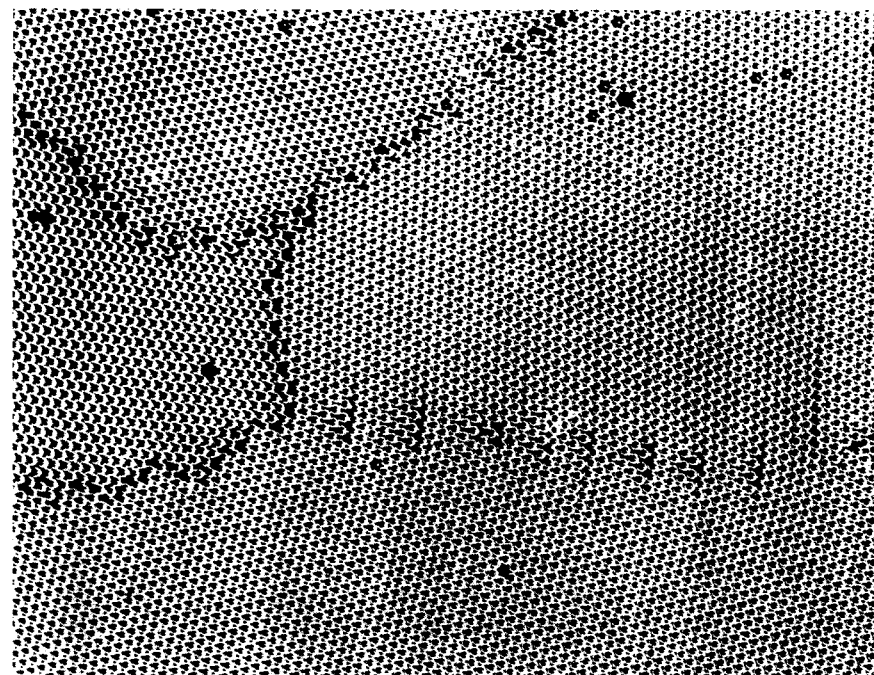


Fig. 3.11 Rafts of soap bubbles showing several grains of varying misorientation. Note that the boundary with the smallest misorientation is made up of a row of dislocations, whereas the high-angle boundaries have a disordered structure in which individual dislocations cannot be identified. (After P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, New York, 1969, from C.S. Smith.)

Measured high-angle grain boundary energies γ_b are often found to be roughly given by

$$\gamma_b \approx \frac{1}{3} \gamma_{sv} \quad (3.11)$$

Some selected values for γ_b and γ_b/γ_{sv} are listed in Table 3.2. As for surface energies γ_b is temperature dependent decreasing somewhat with increasing temperature.

Table 3.2 Measured Grain Boundary Free Energies

Values selected from compilation given in *Interfacial Phenomena in Metals and Alloys*, by L.E. Murr, Addison-Wesley, London, 1975.

Crystal	$\gamma_b/\text{mJ m}^{-2}$	T/°C	γ_b/γ_{sv}
Sn	164	223	0.24
Al	324	450	0.30
Ag	375	950	0.33
Au	378	1000	0.27
Cu	625	925	0.36
γ -Fe	756	1350	0.40
δ -Fe	468	1450	0.23
Pt	660	1300	0.29
W	1080	2000	0.41

3.3.2 Special High-Angle Grain Boundaries

Not all high-angle boundaries have an open disordered structure. There are some *special high-angle boundaries* which have significantly lower energies than the random boundaries. These boundaries only occur at particular misorientations and boundary planes which allow the two adjoining lattices to fit together with relatively little distortion of the interatomic bonds.

The simplest special high-angle grain boundary is the boundary between two twins. If the twin boundary is parallel to the twinning plane the atoms in the boundary fit perfectly into both grains. The result is a *coherent twin boundary* as illustrated in Fig. 3.12a. In fcc metals this is a {111} close-packed plane. Because the atoms in the boundary are essentially in undistorted positions the energy of a coherent twin boundary is extremely low in comparison to the energy of a random high-angle boundary.

If the twin boundary does not lie exactly parallel to the twinning plane, Fig. 3.12b, the atoms do not fit perfectly into each grain and the boundary energy is much higher. This is known as an *incoherent twin boundary*. The energy of a twin boundary is therefore very sensitive to the orientation of the

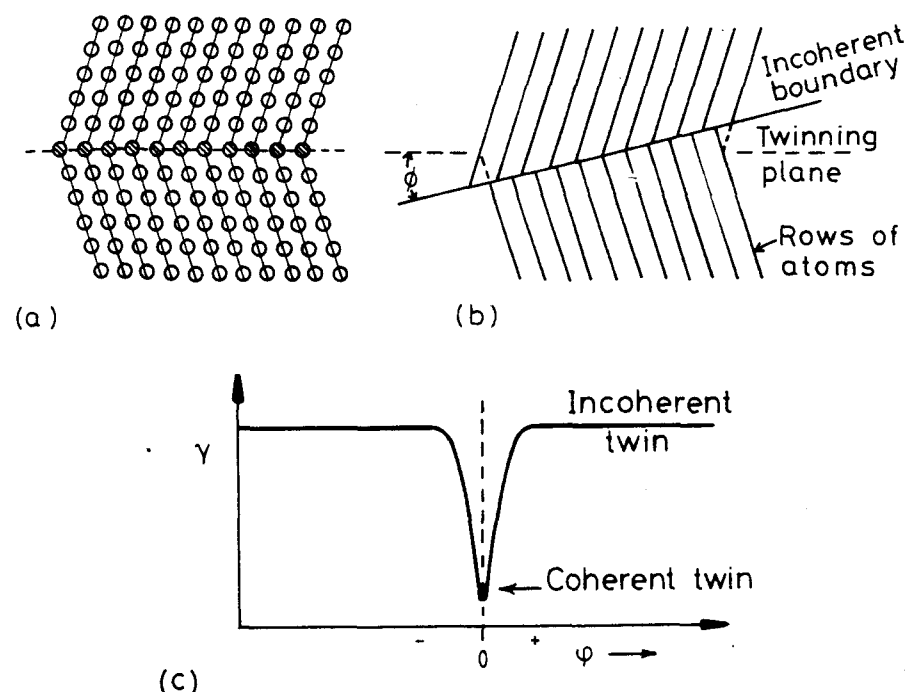


Fig. 3.12 (a) A coherent twin boundary. (b) An incoherent twin boundary. (c) Twin-boundary energy as a function of the grain boundary orientation.

boundary plane. If γ is plotted as a function of the boundary orientation a sharp cusped minimum is obtained at the coherent boundary position as shown in Fig. 3.12c. Table 3.3 lists some experimentally measured values of coherent and incoherent twins along with high-angle grain boundary energies for comparison.

Table 3.3 Measured Boundary Free Energies for Crystals in Twin Relationships (Units mJ m^{-2})

Values selected from compilation given in *Interfacial Phenomena in Metals and Alloys*, by L.E. Murr, Addison-Wesley, London, 1975.

Crystal	Coherent twin boundary energy	Incoherent twin boundary energy	Grain boundary energy
Cu	21	498	623
Ag	8	126	377
Fe-Cr-Ni (stainless steel type 304)	19	209	835

Twin orientations in fcc metals correspond to a misorientation of 70.5° about a $\langle 110 \rangle$ axis. Therefore a twin boundary is a special high-angle grain boundary, and a coherent twin boundary is a symmetrical tilt boundary between the two twin-related crystals. Figure 3.13 shows measured grain-boundary energies for various symmetric tilt boundaries in aluminium. When the two grains are related by a rotation about a $\langle 100 \rangle$ axis, Fig. 3.13a, it can be seen that most high-angle boundaries have about the same energy and should therefore have a relatively disordered structure characteristic of random boundaries. However, when the two grains are related by a rotation about a $\langle 110 \rangle$ axis there are several large-angle orientations which have significantly lower energies than the random boundaries (Fig. 3.13b). $\theta = 70.5^\circ$ corresponds to the coherent twin boundary discussed above, but low-energy boundaries are also found for several other values of θ . The reasons for these other special grain boundaries are not well understood. However, it seems reasonable to suppose that the atomic structure of these boundaries is such that they contain extensive areas of good fit. A two-dimensional example is shown in Fig. 3.14. This is a symmetrical tilt boundary between grains with a misorientation of 38.2° . The boundary atoms fit rather well into both grains leaving relatively little free volume. Moreover, a small group of atoms (shaded) are repeated at regular intervals along the boundary.

3.3.3 Equilibrium in Polycrystalline Materials

Let us now examine how the possibility of different grain-boundary energies affects the microstructure of a polycrystalline material. Figure 3.15 shows the microstructure of an annealed austenitic stainless steel (fcc). The material contains high- and low-angle grain boundaries as well as coherent and incoherent twin boundaries. This microstructure is determined by how the different grain boundaries join together in space. When looking at two-

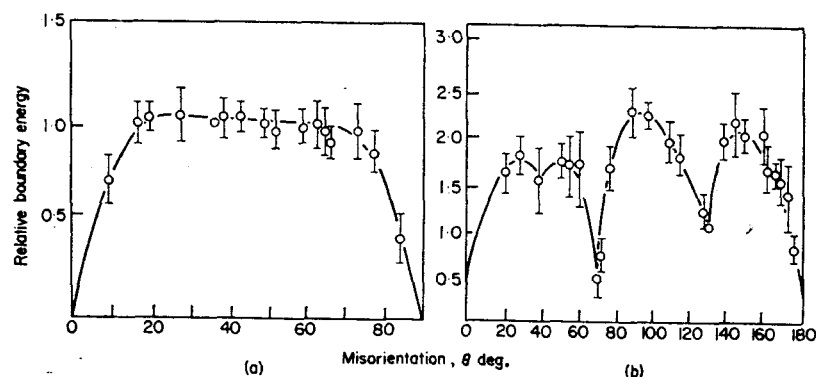


Fig. 3.13 Measured grain boundary energies for symmetric tilt boundaries in Al (a) when the rotation axis is parallel to $\langle 100 \rangle$, (b) when the rotation axis is parallel to $\langle 110 \rangle$. (After G. Hasson and C. Goux, *Scripta Metallurgica*, 5 (1971) 889.)

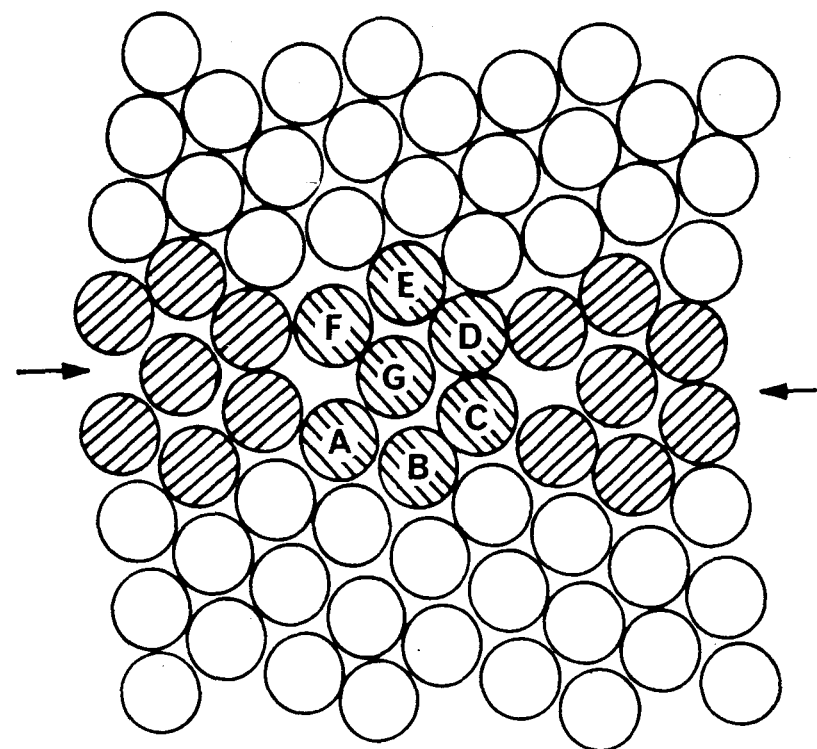


Fig. 3.14 Special grain boundary. (After H. Gleiter, *Physica Status Solidi* (b) 45 (1971) 9.)

dimensional microstructures like this it is important to remember that in reality the grains fill three dimensions, and only one section of the three-dimensional network of internal boundaries is apparent. Note that two grains meet in a plane (a grain boundary) three grains meet in a line (a grain edge) and four grains meet at a point (a grain corner). Let us now consider the factors that control the grain shapes in a recrystallized polycrystal.

The first problem to be solved is why grain boundaries exist at all in annealed materials. The boundaries are all high-energy regions that increase the free energy of a polycrystal relative to a single crystal. Therefore a polycrystalline material is never a true equilibrium structure. However the grain boundaries in a polycrystal can adjust themselves during annealing to produce a *metastable* equilibrium at the grain boundary intersections.

The conditions for equilibrium at a grain-boundary junction can be obtained either by considering the total grain boundary energy associated with a particular configuration or, more simply, by considering the forces that each boundary exerts on the junction. Let us first consider a grain-boundary segment of unit width and length OP as shown in Fig. 3.16. If the boundary is mobile then forces F_x and F_y must act at O and P to maintain the boundary in

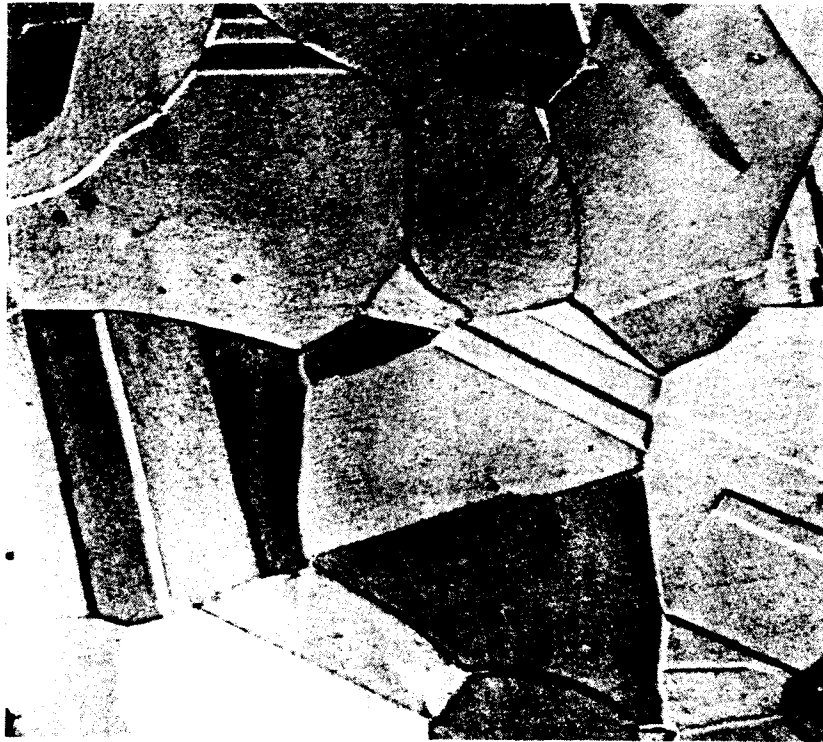


Fig. 3.15 Microstructure of an annealed crystal of austenitic stainless steel. (After P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, New York, 1969.)

equilibrium. From Equation 3.3, $F_x = \gamma$. F_y can be calculated as follows: if P is moved a small distance δy while O remains stationary, the work done will be $F_y \delta y$. This must balance the increase in boundary energy caused by the change in orientation $\delta \theta$, i.e.

$$F_y \delta y = l \frac{d\gamma}{d\theta} \delta \theta$$

Since $\delta y = l \delta \theta$

$$F_y = \frac{d\gamma}{d\theta} \quad (3.12)$$

This means that if the grain-boundary energy is dependent on the orientation of the boundary (Fig. 3.16b) a force $d\gamma/d\theta$ must be applied to the ends of the boundary to prevent it rotating into a lower energy orientation. $d\gamma/d\theta$ is therefore known as a *torque term*. Since the segment OP must be supported by forces F_x and F_y the boundary exerts equal but opposite forces $-F_x$ and $-F_y$ on the ends of the segment which can be junctions with other grain boundaries.

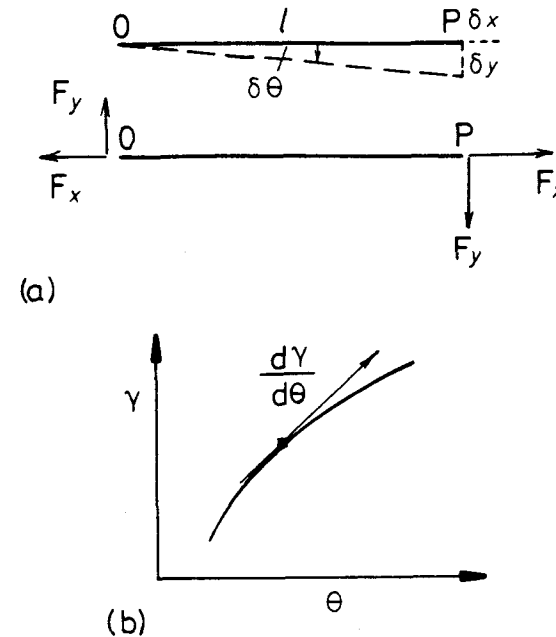


Fig. 3.16 (a) Equilibrium forces F_x and F_y supporting a length l of boundary OP. (b) The origin of F_y .

If the boundary happens to be at the orientation of a cusp in the free energy, e.g. as shown in Fig. 3.12c, there will be no torque acting on the boundary since the energy is a minimum in that orientation. However, the boundary will be able to resist a pulling force F_y of up to $(d\gamma/d\theta)_{\text{cusp}}$ without rotating.

If the boundary energy is independent of orientation the torque term is zero and the grain boundary behaves like a soap film. Under these conditions the requirement for metastable equilibrium at a junction between three grains, Fig. 3.17, is that the boundary tensions γ_1 , γ_2 and γ_3 must balance. In

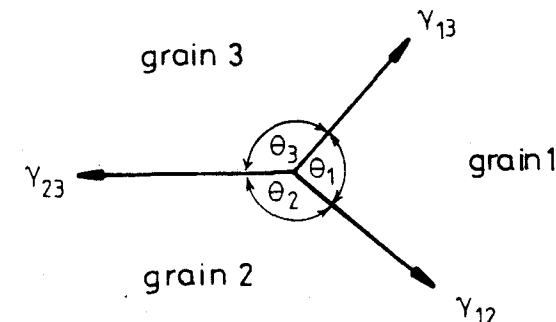


Fig. 3.17 The balance of grain boundary tensions for a grain boundary intersection in metastable equilibrium.

mathematical terms

$$\frac{\gamma_{23}}{\sin \theta_1} = \frac{\gamma_{13}}{\sin \theta_2} = \frac{\gamma_{12}}{\sin \theta_3} \quad (3.13)$$

Equation 3.13 applies to any three boundaries so that grain 1 for example could be a different phase to grains 2 and 3. Alternatively grain 1 could be a vapour phase in which case γ_{13} and γ_{12} would be the surface energies of the solid. This relationship is therefore useful for determining relative boundary energies.

One method of measuring grain-boundary energy is to anneal a specimen at a high temperature and then measure the angle at the intersection of the surface with the boundary, see Fig. 3.18. If the solid-vapour energy (γ_{sv}) is the same for both grains, balancing the interfacial tensions gives

$$2\gamma_{sv} \cos \frac{\theta}{2} = \gamma_b \quad (3.14)$$

Therefore if γ_{sv} is known γ_b can be calculated.

When using Equation 3.14 it must be remembered that the presence of any torque terms has been neglected and such an approximation may introduce large errors. To illustrate the importance of such effects let us consider the junction between coherent and incoherent twin boundary segments, Fig. 3.19. As a result of the orientation dependence of twin boundary energy, Fig. 3.12c, it is energetically favourable for twin boundaries to align themselves parallel to the twinning plane. If, however, the boundary is constrained to follow a *macroscopic plane* that is near but not exactly parallel to the twinning plane the boundary will usually develop a stepped appearance with large coherent low-energy facets and small incoherent high-energy risers as shown in Fig. 3.19. Although this configuration does not minimize the total twin boundary area it does minimize the total free energy.

It is clear that at the coherent/incoherent twin junction the incoherent twin boundary tension γ_i must be balanced by a torque term. Since the maximum value of the resisting force is $d\gamma_c/d\theta$, the condition that the configuration shown in Fig. 3.19 is stable is

$$\gamma_i \leq \frac{d\gamma_c}{d\theta} \quad (3.15)$$

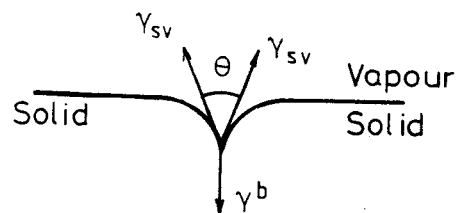


Fig. 3.18 The balance of surface and grain boundary tensions at the intersection of a grain boundary with a free surface.

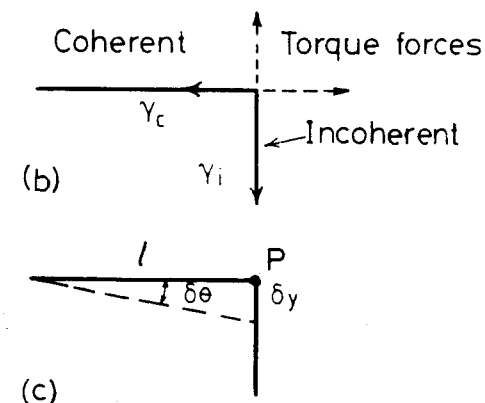
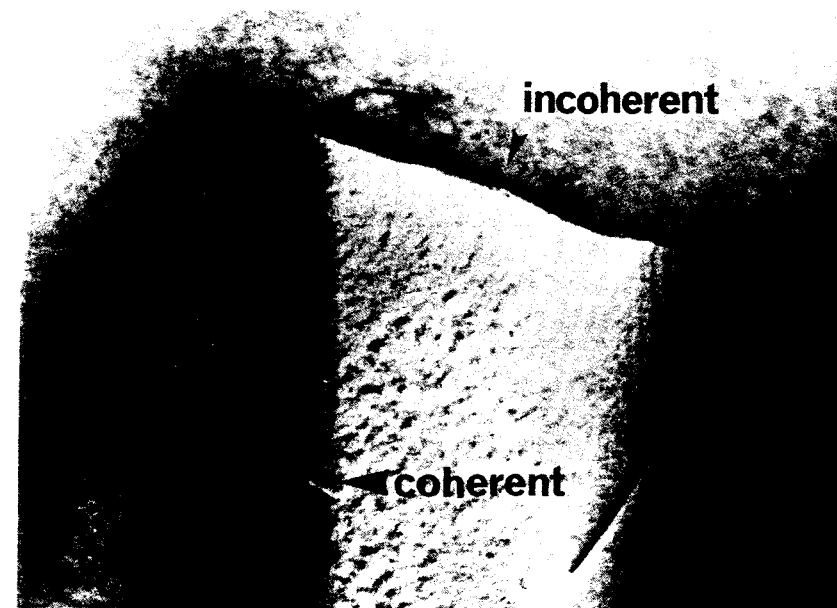


Fig. 3.19 (a) A twin boundary in a thin foil specimen as imaged in the transmission electron microscope. (After M.N. Thompson and C.W. Chen, *Philips Electron Optics Bulletin*, EM 112-1979/1 Eindhoven, 1979.) (b) and (c), the coherent and incoherent segments of the twin boundary.

Likewise the 'incoherent' facet must also be a special boundary showing rather good fit in order to provide a force resisting γ_c . That is

$$\gamma_c \leq \frac{d\gamma_i}{d\theta} \quad (3.16)$$

However, since γ_c is usually very small the incoherent interface need only lie in a rather shallow energy cusp.

The above can be obtained in another way from energy considerations. If (metastable) equilibrium exists at P in Fig. 3.19c, then a small displacement such as that shown should either produce no change, or an increase in the total free energy of the system, i.e.

$$dG \geq 0$$

Considering unit depth a small displacement δy at P will increase the total free energy by an amount

$$dG = l \frac{d\gamma_c}{d\theta} \delta\theta - \gamma_i \delta y \geq 0$$

Since $l\delta\theta = \delta y$ this leads to the same result as given by Equation 3.15

3.3.4 Thermally Activated Migration of Grain Boundaries

In the previous section it was shown that metastable equilibrium at the grain boundary junctions requires certain conditions to be satisfied for the angles at which three boundaries intersect. For simplicity, if all grain boundaries in a polycrystal are assumed to have the same grain-boundary energy independent of boundary orientation, Equation 3.13 predicts that $\theta_1 = \theta_2 = \theta_3 = 120^\circ$. It can be similarly shown that the grain-boundary edges meeting at a corner formed by four grains will make an angle of $109^\circ 28'$. If these, or similar, angular conditions are satisfied then metastable equilibrium can be established at all grain boundary junctions. However, for a grain structure to be in complete metastable equilibrium the surface tensions must also balance over all the boundary faces between the junctions. If a boundary is curved in the shape of a cylinder, Fig. 3.20a, it is acted on by a force of magnitude γ/r towards its centre of curvature. Therefore the only way the boundary tension forces can balance in three dimensions is if the boundary is planar ($r = \infty$) or if it is curved with equal radii in opposite directions, Fig. 3.20b and c. It is

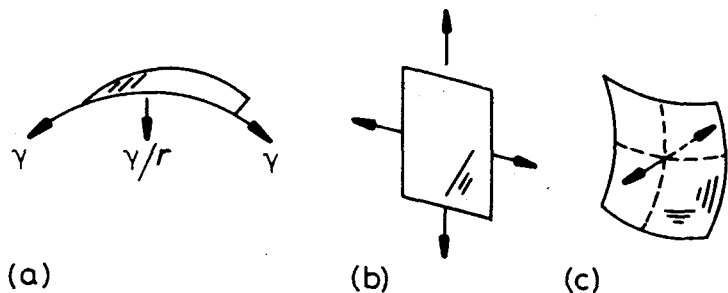


Fig. 3.20 (a) A cylindrical boundary with a radius of curvature r is acted on by a force γ/r . (b) A planar boundary with no net force. (c) A doubly curved boundary with no net force.

theoretically possible to construct a three-dimensional polycrystal in which the boundary tension forces balance at all faces and junctions, but in a random polycrystalline aggregate, typical of real metallurgical specimens, there are always boundaries with a net curvature in one direction. Consequently a random grain structure is inherently unstable and, on annealing at high temperatures, the unbalanced forces will cause the boundaries to migrate towards their centres of curvature.

The effect of different boundary curvatures in *two dimensions* is shown in Fig. 3.21. Again for simplicity it has been assumed that equilibrium at each boundary junction results in angles of 120° . Therefore if a grain has six boundaries they will be planar and the structure metastable. However, if the total number of boundaries around a grain is less than six each boundary must be concave inwards, Fig. 3.21. These grains will therefore shrink and eventually disappear during annealing. Larger grains, on the other hand, will have

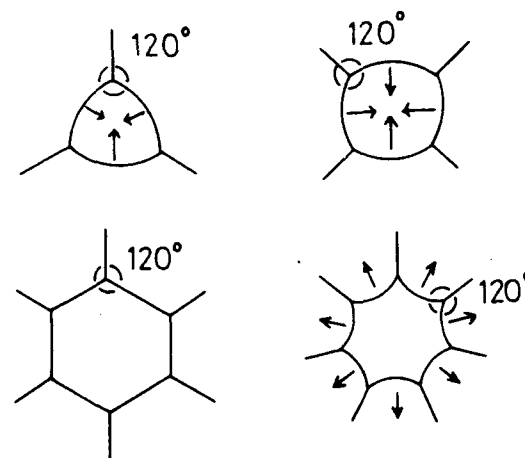


Fig. 3.21 Two-dimensional grain boundary configurations. The arrows indicate the directions boundaries will migrate during grain growth.

more than six boundaries and will grow. The overall result of such boundary migration is to reduce the number of grains, thereby increasing the mean grain size and reducing the total grain boundary energy. This phenomenon is known as *grain growth* or *grain coarsening*. It occurs in metals at temperatures above about $0.5 T_m$ where the boundaries have significant mobility. A soap froth serves as a convenient analogue to demonstrate grain growth as shown in Fig. 3.22.

In the case of the cells in a soap froth the higher pressure on the concave side of the films induces the air molecules in the smaller cells to diffuse through the film into the larger cells, so that the small cells eventually disappear. A similar effect occurs in metal grains. In this case the atoms in the shrinking grain detach themselves from the lattice on the high pressure side of the boundary and relocate themselves on a lattice site of the growing grain.

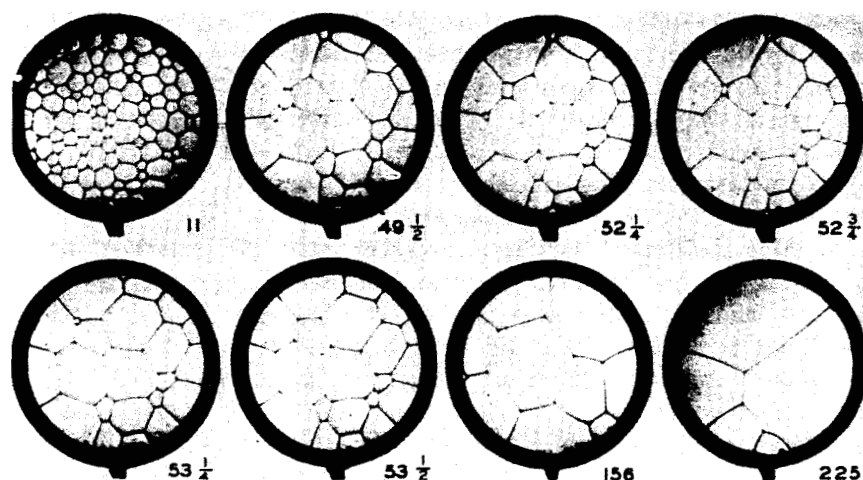


Fig. 3.22 Two-dimensional cells of a soap solution illustrating the process of grain growth. Numbers are time in minutes. (After C.S. Smith, *Metal Interfaces*, American Society for Metals, 1952, p. 81.)

For example in Fig. 3.23a if atom C jumps from grain 1 to grain 2 the boundary locally advances a small distance.

The effect of the pressure difference caused by a curved boundary is to create a difference in *free energy* (ΔG) or *chemical potential* ($\Delta\mu$) that drives the atoms across the boundary, see Fig. 3.24. In a pure metal ΔG and $\Delta\mu$ are identical and are given by Equation 1.58 as

$$\Delta G = \frac{2\gamma V_m}{r} = \Delta\mu \quad (3.17)$$

This free energy difference can be thought of as a *force* pulling the grain boundary towards the grain with the higher free energy. As shown in Fig. 3.25, if unit area of grain boundary advances a distance δx the number of moles of material that enter grain B is $\delta x \cdot 1/V_m$ and the free energy released is given by

$$\Delta G \cdot \delta x / V_m$$

This can be equated to the work done by the pulling force $F\delta x$. Thus the pulling force per unit area of boundary is given by

$$F = \frac{\Delta G}{V_m} \quad \text{N m}^{-2} \quad (3.18)$$

In other words the force on the boundary is simply the free energy difference *per unit volume* of material.

In the case of grain growth ΔG arises from the boundary curvature, but Equation 3.18 applies equally to *any* boundary whose migration causes a

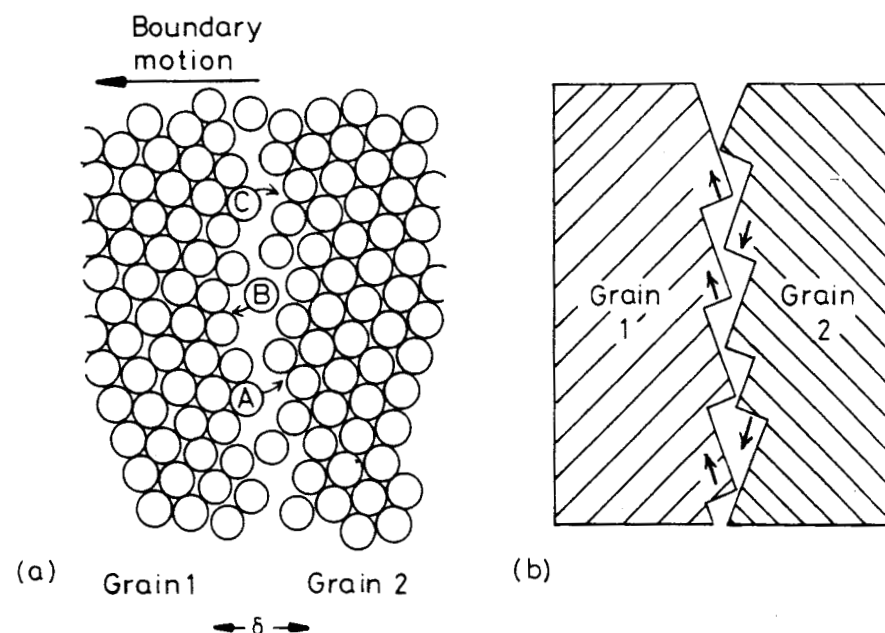


Fig. 3.23 (a) The atomic mechanism of boundary migration. The boundary migrates to the left if the jump rate from grain 1 $\rightarrow$ 2 is greater than 2 $\rightarrow$ 1. Note that the free volume within the boundary has been exaggerated for clarity. (b) Step-like structure where close-packed planes protrude into the boundary.

decrease in free energy. During recrystallization, for example, the boundaries between the new strain-free grains and the original deformed grains are acted on by a force $\Delta G/V_m$ where, in this case, ΔG is due to the difference in dislocation strain energy between the two grains. Figure 3.26 shows a dislocation-free recrystallized grain expanding into the heavily deformed surroundings. In this case the total grain-boundary area is increasing, therefore the

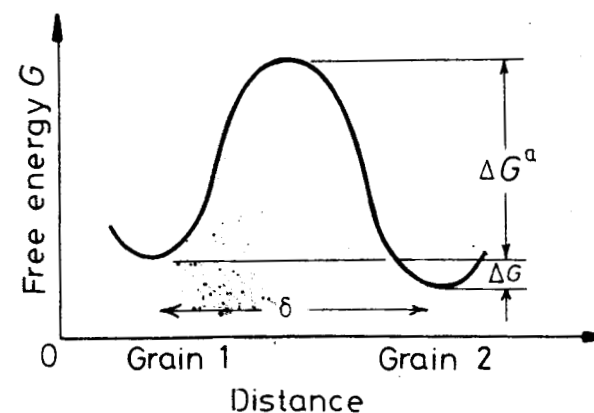


Fig. 3.24 The free energy of an atom during the process of jumping from one grain to the other.

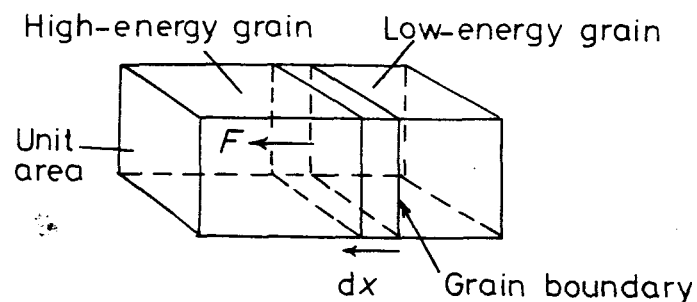


Fig. 3.25 A boundary separating grains with different free energies is subjected to a pulling force F .



Fig. 3.26 Grain boundary migration in nickel pulled 10% and annealed 10 min at 425 °C. The region behind the advancing boundary is dislocation-free. (After J. Bailey and P. Hirsch, *Proceedings of the Royal Society, London*, A267 (1962) 11.)

driving force for recrystallization must be greater than the opposing boundary tension forces. Such forces are greatest when the new grain is smallest, and the effect is therefore important in the early stages of recrystallization.

Let us now consider the effect of the driving force on the kinetics of boundary migration. In order for an atom to be able to break away from grain 1 it must acquire, by thermal activation, an activation energy ΔG^a , Fig 3.24. If the atoms vibrate with a frequency ν_1 the number of times per second that an atom has this energy is $\nu_1 \exp(-\Delta G^a/RT)$. If there are on average n_1 atoms per unit area in a favourable position to make a jump there will be $n_1 \nu_1 \exp(-\Delta G^a/RT)$ jumps $\text{m}^{-2} \text{s}^{-1}$ away from grain 1. It is possible that not all these atoms will find a suitable site and 'stick' to grain 2. If the probability of being accommodated in grain 2 is A_2 the effective flux of atoms from grain 1 to 2 will be

$$A_2 n_1 \nu_1 \exp(-\Delta G^a/RT) \text{ m}^{-2} \text{s}^{-1}$$

There will also be a similar flux in the reverse direction, but if the atoms in grain 2 have a lower free energy than the atoms in grain 1 by ΔG (mol^{-1}) the flux from 2 to 1 will be

$$A_1 n_2 \nu_2 \exp(-(\Delta G^a + \Delta G)/RT) \text{ m}^{-2} \text{s}^{-1}$$

When $\Delta G = 0$ the two grains are in equilibrium and there should therefore be no net boundary movement, i.e. the rates at which atoms cross the boundary in opposite directions must be equal. Equating the above expressions then gives

$$A_1 n_2 \nu_2 = A_2 n_1 \nu_1$$

For a high-angle grain boundary it seems reasonable to expect that there will not be great problems with accommodation so that $A_1 \approx A_2 \approx 1$. Assuming the above equality also holds for small non-zero driving forces, with $\Delta G > 0$ there will be a net flux from grain 1 to 2 given by

$$J_{\text{net}} = A_2 n_1 \nu_1 \exp\left(-\frac{\Delta G^a}{RT}\right) \left\{1 - \exp\left(-\frac{\Delta G}{RT}\right)\right\} \quad (3.19)$$

If the boundary is moving with a velocity v the above flux must also be equal to $v/(V_m/N_a)$, where (V_m/N_a) is the atomic volume. Therefore expanding $\exp(-\Delta G/RT)$ for the usual case of $\Delta G \ll RT$ gives

$$v = \frac{A_2 n_1 \nu_1 V_m^2}{N_a RT} \exp\left(-\frac{\Delta G^a}{RT}\right) \frac{\Delta G}{V_m} \quad (3.20)$$

In other words v should be proportional to the driving force $\Delta G/V_m$ (N m^{-2}). Equation 3.20 can be written more simply as

$$v = M \cdot \Delta G/V_m \quad (3.21)$$

where M is the mobility of the boundary, i.e. the velocity under unit driving force. Substituting for ΔG^a gives

$$M = \left\{ \frac{A_2 n_1 v_1 V_m^2}{N_a R T} \exp \left(\frac{\Delta S^a}{R} \right) \right\} \exp \left(\frac{-\Delta H^a}{RT} \right) \quad (3.22)$$

Note how this simple model predicts an exponential increase in mobility with temperature. This result should of course be intuitively obvious since the boundary migration is a thermally activated process like diffusion. Indeed boundary migration and boundary diffusion are closely related processes. The only difference is that diffusion involves transport along the boundary whereas migration requires atomic movement across the boundary.

The model used to derive Equations 3.20 and 3.22 is particularly simple and gross assumptions are involved. In real grain boundaries it is likely that not all atoms in the boundary are equivalent and some will jump more easily than others. For example atoms may jump preferably to and from atomic steps or ledges, like atoms A, B and C in Fig. 3.23a. In fcc metals such ledges should exist where the close-packed {111} planes protrude into the boundary. Boundary migration could then be effected by the growth of the ledges in one grain combined with the shrinking of corresponding ledges in the other grain as shown in Fig. 3.23b.

From our discussion of grain-boundary structure it might be argued that the relatively open structure of a random high-angle boundary should lead to a high mobility whereas the denser packing of the special boundaries should be associated with a low mobility. Indeed, the coherent twin boundary, in which the atoms fit perfectly into both grains, has been found to be almost entirely immobile⁶. However, experiments have shown that the other special boundaries are usually more mobile than random high-angle boundaries. The reason for this is associated with the presence of impurity or alloying elements in the metal. Figure 3.27 shows data for the migration of various boundaries in zone-refined lead alloyed with different concentrations of tin. For a given driving force the velocity of the random boundaries decreases rapidly with increasing alloy content. Note that only very low concentrations of impurity are required to change the boundary mobility by orders of magnitude. The special grain boundaries on the other hand are less sensitive to impurities. It is possible that if the metal were 'perfectly' pure the random boundaries would have the higher mobility. The reason for this type of behaviour arises from differences in the interactions of alloy elements or impurities with different boundaries.

Generally the grain boundary energy of a pure metal changes on alloying. Often (though not always) it is reduced. Under these circumstances the concentration of alloying element in the boundary is higher than that in the matrix. In grain boundary segregation theory, grain boundary solute concentrations (X_b) are expressed as fractions of a monolayer. One monolayer ($X_b = 1$) means that the solute atoms in the boundary could be arranged to form a single close-packed layer of atoms. Approximately, for low mole fractions of solute in the matrix (X_0), the boundary solute concentration X_b is given by

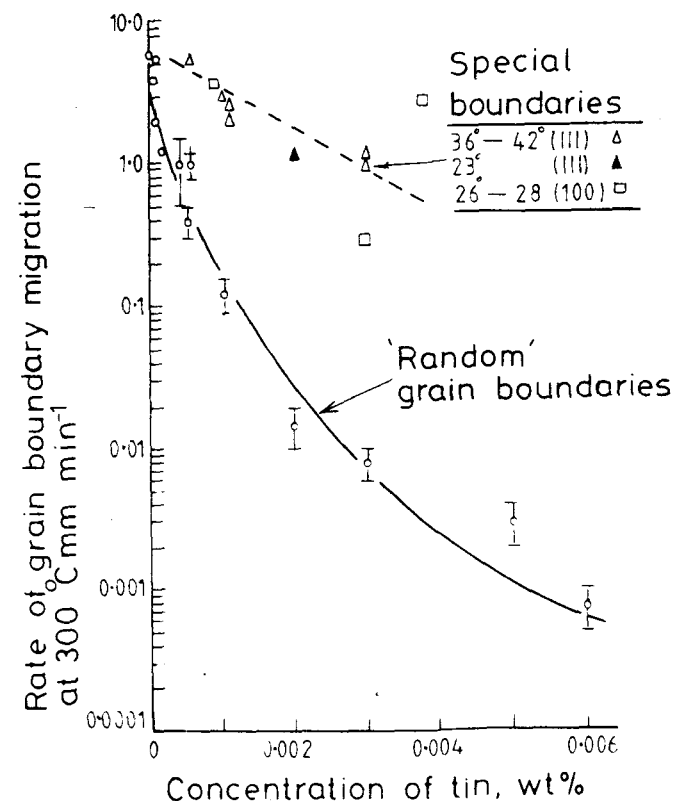


Fig. 3.27 Migration rates of special and random boundaries at 300 °C in zone-refined lead alloyed with tin under equal driving forces. (After K. Aust and J.W. Rutter, *Transactions of the Metallurgical Society of AIME*, 215 (1959) 119.)

$$X_b \approx X_0 \exp \frac{\Delta G_b}{RT} \quad (3.23)$$

ΔG_b is the free energy released per mole when a solute atom is moved from the matrix to the boundary. ΔG_b is usually positive and roughly increases as the size misfit between the solute and matrix increases and as the solute-solute bond strength decreases.

Equation 3.23 shows how grain boundary segregation decreases as temperature increases, i.e. the solute 'evaporates' into the matrix. For sufficiently low temperatures or high values of ΔG_b , X_b increases towards unity and Equation 3.23 breaks down as X_b approaches a maximum saturation value.

The variation of boundary mobility with alloy concentration varies markedly from one element to another. It is a general rule that ΔG_b , which measures the tendency for segregation, increases as the matrix solubility decreases. This is illustrated by the experimental data in Fig. 3.28.

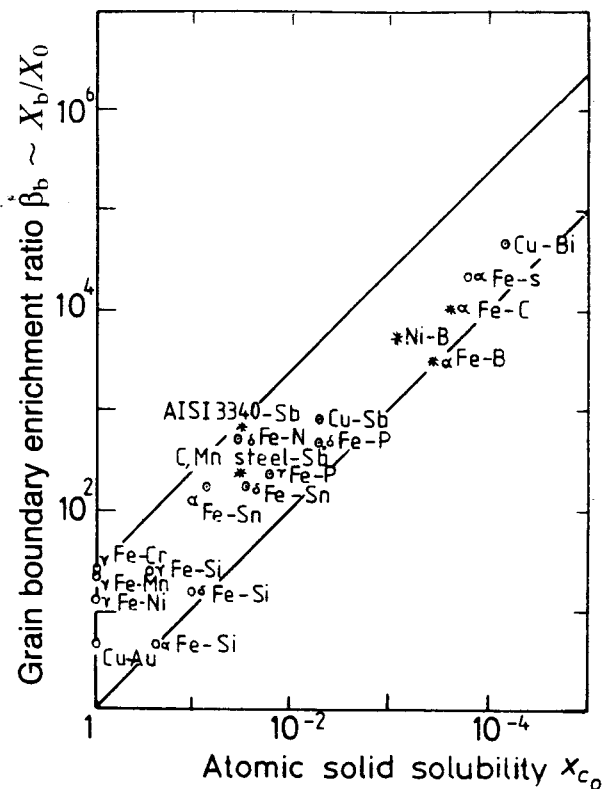


Fig. 3.28 Increasing grain boundary enrichment with decreasing solid solubility in a range of systems. (After E.D. Hondras and M.P. Seah, *International Metallurgical Reviews*, December 1977, Review 222.)

When the boundary moves the solute atoms migrate along with the boundary and exert a *drag* that reduces the boundary velocity. The magnitude of the drag will depend on the binding energy and the concentration in the boundary. The higher mobility of special boundaries can, therefore, possibly be attributed to a low solute drag on account of the relatively more close-packed structure of the special boundaries.

The variation of boundary mobility with alloy concentration varies markedly from one element to another. It is a general rule that Q_B , which measures the tendency for segregation, increases as the matrix solubility decreases. This is illustrated by the experimental data in Fig. 3.28.

It is possible that the higher mobility of special grain boundaries plays a role in the development of recrystallization textures. If a polycrystalline metal is heavily deformed, by say rolling to a 90% reduction, a deformation texture develops such that the rolled material resembles a deformed single crystal. On heating to a sufficiently high temperature new grains nucleate and begin to grow. However, not all grains will grow at the same rate: those grains which

are specially oriented with respect to the matrix should have higher mobility boundaries and should overgrow the boundaries of the randomly oriented grains. Consequently the recrystallized structure should have a special orientation with respect to the original 'single crystal'. Thus a new texture results which is called the recrystallization texture. Recrystallization is, however, incompletely understood and the above explanation of recrystallization texture may be an oversimplification. It is possible for example that the nuclei for recrystallization are themselves specially oriented with respect to the deformed matrix.

A recrystallization texture is sometimes an advantage. For example the proper texture in Fe-3wt% Si alloys makes them much better soft magnets for use in transformers. Another application is in the production of textured sheet for the deep drawing of such materials as low-carbon steel. The only way to avoid a recrystallization texture is to give an intermediate anneal before a deformation texture has been produced.

3.3.5 The Kinetics of Grain Growth

It was shown in the previous section that at sufficiently high temperatures the grain boundaries in a recrystallized specimen will migrate so as to reduce the total number of grains and thereby increase the mean grain diameter. In a single-phase metal the rate at which the mean grain diameter $\bar{D}$ increases with time will depend on the grain boundary mobility and the driving force for boundary migration.

If we assume that the mean radius of curvature of all the grain boundaries is proportional to the mean grain diameter $\bar{D}$ the mean driving force for grain growth will be proportional to $2\gamma/\bar{D}$ (Equation 3.17). Therefore

$$\bar{v} = \alpha M \frac{2\gamma}{\bar{D}} = \frac{d\bar{D}}{dt} \quad (3.24)$$

where α is a proportionality constant of the order of unity.

Note that this equation implies that the rate of grain growth is inversely proportional to $\bar{D}$ and increases rapidly with increasing temperature due to increased boundary mobility, M . Integration of Equation 3.24 taking $\bar{D} = D_0$ when $t = 0$ gives

$$\bar{D}^2 = D_0^2 + Kt \quad (3.25)$$

where $K = 4\alpha M\gamma$.

Experimentally it is found that grain growth in single-phase metals follows a relationship of the form

$$\bar{D} = K't^n \quad (3.26)$$



Fig. 3.29 Optical micrograph ($\times 130$) showing abnormal grain growth in a fine-grained steel containing 0.4 wt% carbon. The matrix grains are prevented from growing by a fine dispersion of unresolved carbide particles. (After D.T. Gawne and G.T. Higgins, *Journal of the Iron and Steel Institute*, 209 (1971), 562.)

where K' is a proportionality constant which increases with temperature. This is equivalent to Equation 3.25 with $n = 0.5$ if $D \geq D_0$. However, the experimentally determined values of n are usually much less than 0.5 and only approach 0.5 in very pure metals or at very high temperatures. The reasons for this are not fully understood, but the most likely explanation is that the velocity of grain boundary migration, v , is not a linear function of the driving force, ΔG , i.e. the mobility in Equation 3.21 is not a constant but varies with ΔG and therefore also with D . It has been suggested that such a variation of M could arise from solute drag effects⁷.

The above type of grain growth is referred to as *normal*. Occasionally so-called *abnormal* grain growth can occur. This situation is characterized by the growth of just a few grains to very large diameters. These grains then expand consuming the surrounding grains, until the fine grains are entirely replaced by a coarse-grained array. This effect is illustrated in Fig. 3.29 and is also known as discontinuous grain growth, coarsening, or secondary recrystallization. It can occur when normal grain growth ceases due to the presence of a fine precipitate array.

The nature of normal grain growth in the presence of a second phase deserves special consideration. The moving boundaries will be attached to

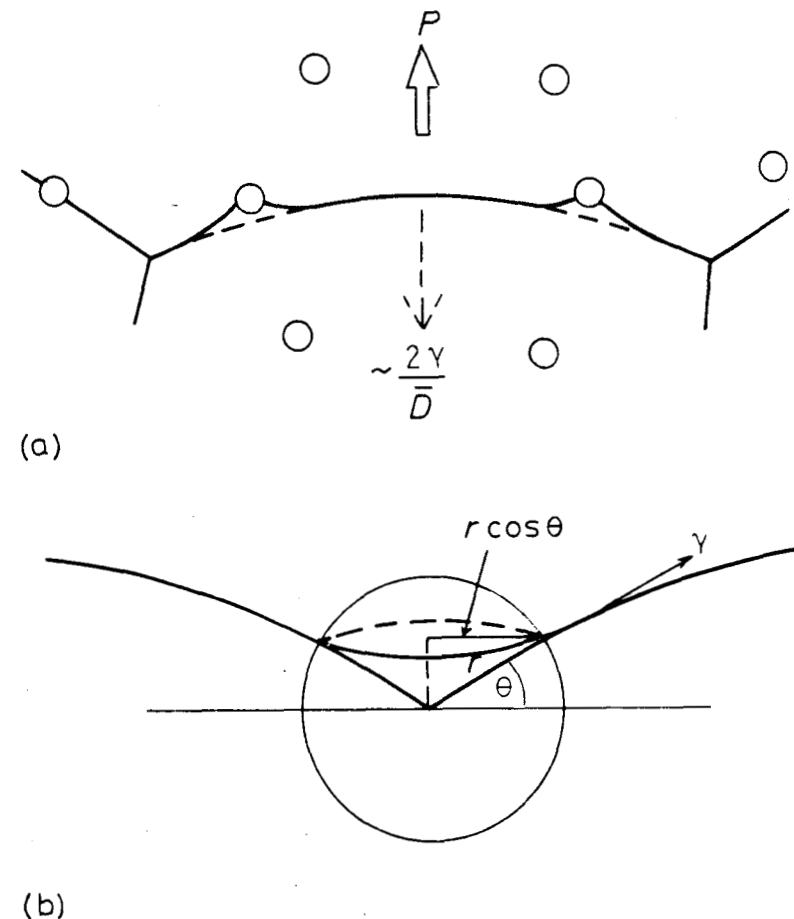


Fig. 3.30 The effect of spherical particles on grain boundary migration.

the particles as shown in Fig. 3.30a, so that the particles exert a pulling force on the boundary restricting its motion. The boundary shown in Fig. 3.30b will be attached to the particle along a length $2\pi r \cos \theta$. Therefore if the boundary intersects the particle surface at 90° the particle will feel a pull of $(2\pi r \cdot \cos \theta \cdot \gamma) \sin \theta$. This will be counterbalanced by an equal and opposite force acting on the boundary. As the boundary moves over the particle surface θ changes and the drag reaches a maximum value when $\sin \theta \cdot \cos \theta$ is a maximum, i.e. at $\theta = 45^\circ$. The maximum force exerted by a single particle is therefore given by $\pi r \gamma$.

If there is a volume fraction f of particles all with a radius r the mean number of particles intersecting unit area of a random plane is $3f/2\pi r^2$ so that the restraining force per unit area of boundary is approximately

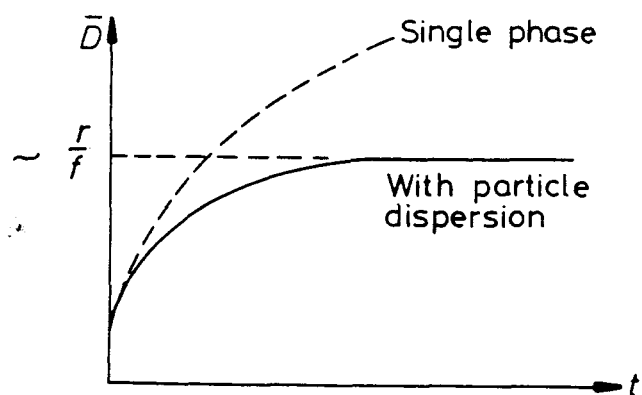


Fig. 3.31 Effect of second-phase particles on grain growth.

$$P = \frac{3f}{2\pi r^2} \cdot \pi r \gamma = \frac{3f\gamma}{2r} \quad (3.27)$$

This force will oppose the driving force for grain growth, namely $\sim 2\gamma/\bar{D}$ as shown in Fig. 3.30a. When $\bar{D}$ is small P will be relatively insignificant, but as $\bar{D}$ increases the driving force $2\gamma/\bar{D}$ decreases and when

$$\frac{2\gamma}{\bar{D}} = \frac{3f\gamma}{2r}$$

the driving force will be insufficient to overcome the drag of the particles and grain growth stagnates. A maximum grain size will be given by

$$\bar{D}_{\max} = \frac{4r}{3f} \quad (3.28)$$

The effect of a particle dispersion on grain growth is illustrated in Fig. 3.31. It can be seen that the stabilization of a fine grain size during heating at high temperatures requires a large volume fraction of very small particles. Unfortunately, if the temperature is too high, the particles tend to coarsen or dissolve. When this occurs some boundaries can break away before the others and abnormal grain growth occurs, transforming the fine-grain array into a very coarse-grain structure. For example aluminium-killed steels contain aluminium nitride precipitates which stabilize the austenite grain size during heating. However, their effectiveness disappears above about 1000 °C when the aluminium nitride precipitates start to dissolve.

3.4 Interphase Interfaces in Solids

The previous section dealt in some detail with the structure and properties of boundaries between crystals of the same solid phase. In this section we will be

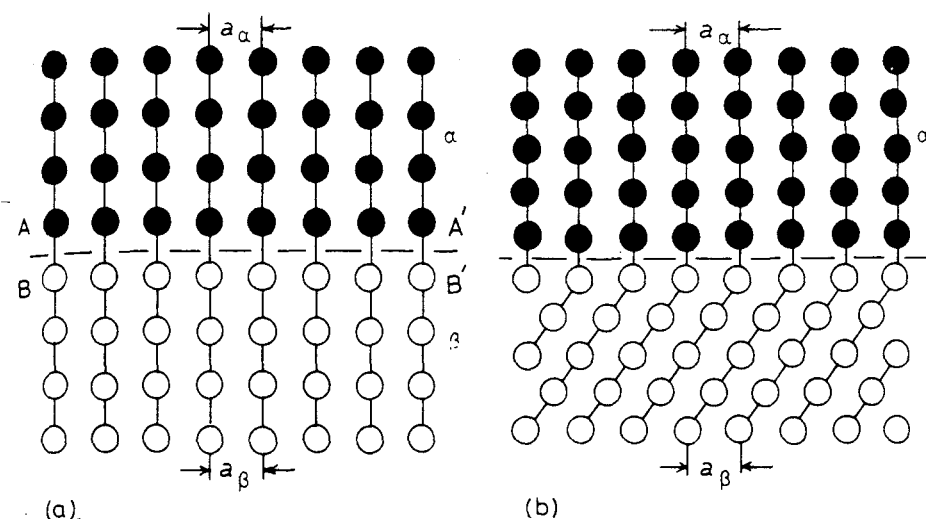


Fig. 3.32 Strain-free coherent interfaces. (a) Each crystal has a different chemical composition but the same crystal structure. (b) The two phases have different lattices.

dealing with boundaries between different solid phases, i.e. where the two adjoining crystals can have different crystal structures and/or compositions. Interphase boundaries in solids can be divided on the basis of their atomic structure into three classes: *coherent*, *semicoherent* and *incoherent*.

3.4.1 Interface Coherence

Fully Coherent Interfaces

A coherent interface arises when the two crystals match perfectly at the interface plane so that the two lattices are continuous across the interface,

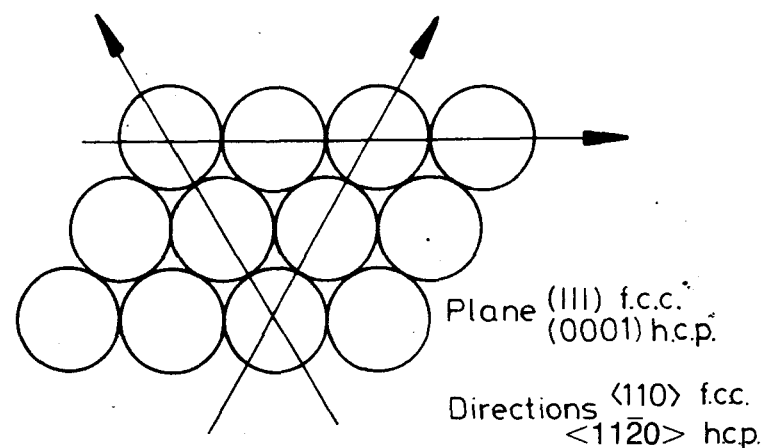


Fig. 3.33 The close-packed plane and directions in fcc and hcp structures.

Fig. 3.32. This can only be achieved if, disregarding chemical species, the *interfacial plane has the same atomic configuration in both phases*, and this requires the two crystals to be oriented relative to each other in a special way. For example such an interface is formed between the hcp silicon-rich κ phase and the fcc copper-rich α -matrix in Cu-Si alloys. The lattice parameters of these two phases are such that the $(111)_{\text{fcc}}$ plane is identical to the $(0001)_{\text{hcp}}$ plane. Both planes are hexagonally close-packed (Fig. 3.33) and in this particular case the interatomic distances are also identical. Therefore when the two crystals are joined along their close-packed planes with the close-packed directions parallel the resultant interface is completely coherent. The requirement that the close-packed planes and directions are parallel produces an *orientation relationship* between the two phases such that

$$\begin{aligned} (111)_{\alpha} // (0001)_{\kappa} \\ [\bar{1}10]_{\alpha} // [11\bar{2}0]_{\kappa} \end{aligned}$$

Note that the relative orientation of two crystals can always be specified by giving two parallel planes (hkl) and two parallel directions [uvw] that lie in those planes.

Within the bulk of each phase every atom has an optimum arrangement of nearest neighbours that produces a low energy. At the interface, however, there is usually a change in composition so that each atom is partly bonded to *wrong* neighbours across the interface. This increases the energy of the interfacial atoms and leads to a chemical contribution to the interfacial energy (γ_{ch}). For a coherent interface this is the only contribution, i.e.

$$\gamma(\text{coherent}) = \gamma_{\text{ch}} \quad (3.29)$$

In the case of the α - κ interface in Cu-Si alloys the interfacial energy has been estimated to be as low as 1 mJ m^{-2} . In general coherent interfacial energies range up to about 200 mJ m^{-2} .

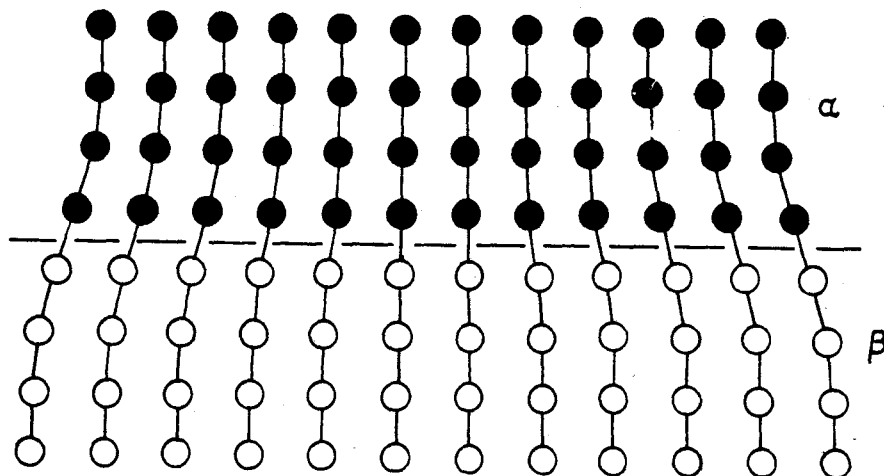


Fig. 3.34 A coherent interface with slight mismatch leads to coherency strains in the adjoining lattices.

In the case of a hcp/fcc interface there is only one plane that can form a coherent interface: no other plane is identical in both crystal lattices. If, however, the two adjoining phases have the same crystal structure and lattice parameter then, apart from differences in composition, all lattice planes are identical.

When the distance between the atoms in the interface is not identical it is still possible to maintain coherency by straining one or both of the two lattices as illustrated in Fig. 3.34. The resultant lattice distortions are known as *coherency strains*.

Semicoherent Interfaces

The strains associated with a coherent interface raise the total energy of the system, and for sufficiently large atomic misfit, or interfacial area, it becomes energetically more favourable to replace the coherent interface with a *semi-coherent* interface in which the disregistry is periodically taken up by *misfit dislocations*, Fig. 3.35.

If d_{α} and d_{β} are the unstressed interplanar spacings of matching planes in the α and β phases respectively, the disregistry, or misfit between the two lattices (δ) is defined by

$$\delta = \frac{d_{\beta} - d_{\alpha}}{d_{\alpha}} \quad (3.30)$$

It can be shown that in one dimension the lattice misfit can be completely accommodated without any long-range strain fields by a set of edge dislocations with a spacing D given by

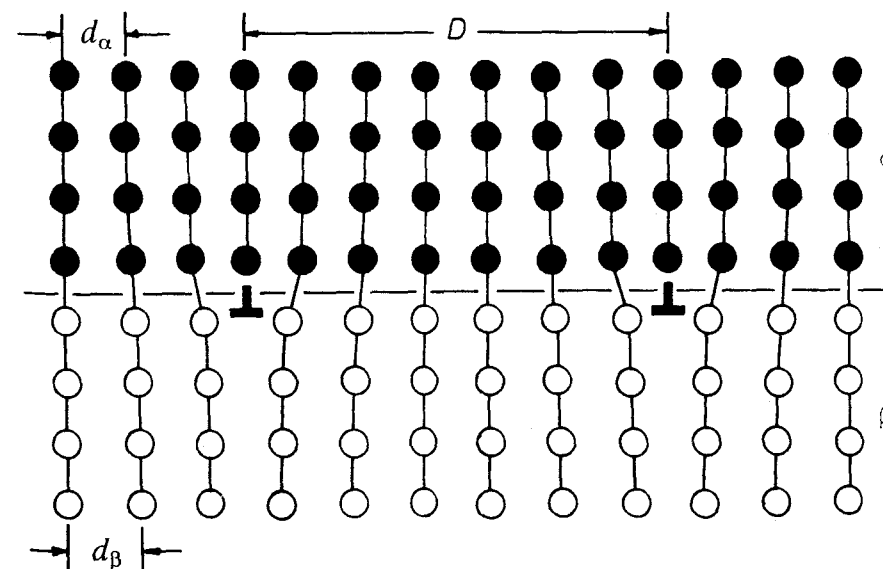


Fig. 3.35 A semicoherent interface. The misfit parallel to the interface is accommodated by a series of edge dislocations.

$$D = \frac{d_\beta}{\delta} \quad (3.31)$$

or approximately, for small δ

$$D = \frac{b}{\delta} \quad (3.32)$$

where $b = (d_\alpha + d_\beta)/2$ is the Burgers vector of the dislocations. The matching in the interface is now almost perfect except around the dislocation cores where the structure is highly distorted and the lattice planes are discontinuous.

In practice misfit usually exists in two dimensions and in this case the coherency strain fields can be completely relieved if the interface contains two non-parallel sets of dislocations with spacings $D_1 = b_1/\delta_1$ and $D_2 = b_2/\delta_2$, as shown in Fig. 3.36. If, for some reason, the dislocation spacing is greater than given by Equation 3.32, the coherency strains will have been only partially relieved by the misfit dislocations and residual long-range strain fields will still be present.

The interfacial energy of a semicoherent interface can be approximately considered as the sum of two parts: (a) a chemical contribution, γ_{ch} , as for a fully coherent interface, and (b) a structural term γ_{st} , which is the extra energy due to the structural distortions caused by the misfit dislocations, i.e.

$$\gamma \text{ (semicoherent)} = \gamma_{ch} + \gamma_{st} \quad (3.33)$$

Equation 3.32 shows that as the misfit δ increases the dislocation spacing diminishes. For *small values of* δ the structural contribution to the interfacial energy is approximately proportional to the density of dislocations in the interface, i.e.

$$\gamma_{st} \propto \delta \quad (\text{for small } \delta) \quad (3.34)$$

However γ_{st} increases less rapidly as δ becomes larger and it levels out when $\delta \approx 0.25$ in a similar way to the variation of grain-boundary energy with θ

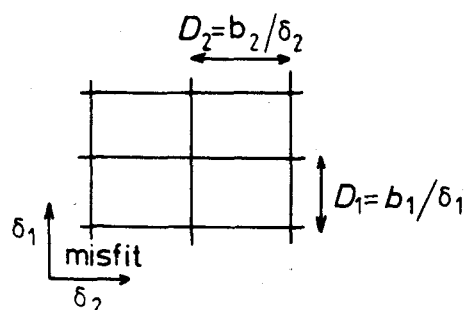


Fig. 3.36 Misfit in two directions (δ_1 and δ_2) can be accommodated by a cross-grid of edge dislocations with spacings $D_1 = b_1/\delta_1$ and $D_2 = b_2/\delta_2$.

shown in Fig. 3.9. The reason for such behaviour is that as the misfit dislocation spacing decreases the associated strain fields increasingly overlap and annul each other. The energies of semicoherent interfaces are generally in the range $200\text{--}500 \text{ mJ m}^{-2}$.

When $\delta > 0.25$, i.e. one dislocation every four interplanar spacings, the regions of poor fit around the dislocation cores overlap and the interface cannot be considered as coherent, i.e. it is *incoherent*.

Incoherent Interfaces

When the interfacial plane has a very different atomic configuration in the two adjoining phases there is no possibility of good matching across the interface. The pattern of atoms may either be very different in the two phases or, if it is similar, the interatomic distances may differ by more than 25%. In both cases the interface is said to be incoherent. In general, incoherent interfaces result when two randomly-oriented crystals are joined across any interfacial plane as shown in Fig. 3.37. They may, however, also exist between crystals with an orientation relationship if the interface has a different structure in the two crystals.

Very little is known about the detailed atomic structure of incoherent interfaces, but they have many features in common with high-angle grain boundaries. For example they are characterized by a high energy ($\sim 500\text{--}1000 \text{ mJ m}^{-2}$) which is relatively insensitive to the orientation of the interfa-

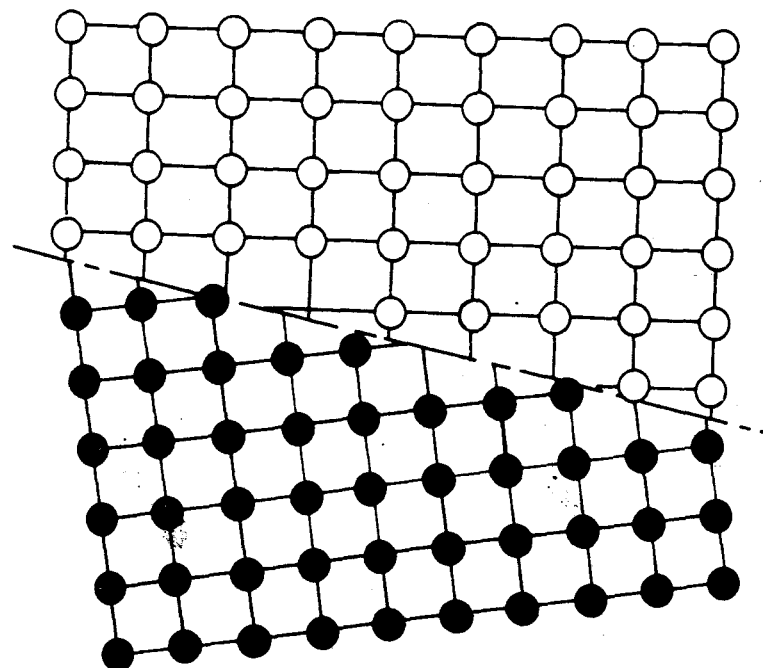


Fig. 3.37 An incoherent interface.

cial plane. They probably have a disordered atomic structure in that the interface lacks the long-range periodicity of coherent and semicoherent interfaces; although, like high-angle grain boundaries, they may have a step-like structure caused by low-index planes protruding into the interface, as in Fig. 3.23b.

Complex Semicoherent Interfaces

The semicoherent interfaces considered above have been observed at boundaries formed by low-index planes whose atom patterns and spacings are clearly almost the same. However, semicoherent interfaces, i.e. interfaces containing misfit dislocations, can also form between phases when good lattice matching is not initially obvious. For example, fcc and bcc crystals often appear with the closest-packed planes in each phase, $(111)_{\text{fcc}}$ and $(110)_{\text{bcc}}$, almost parallel to each other. Two variants of this relationship are found: the so-called Nishiyama–Wasserman (N–W) relationship:

$$(110)_{\text{bcc}} // (111)_{\text{fcc}}, [001]_{\text{bcc}} // [\bar{1}01]_{\text{fcc}}$$

and the so-called Kurdjumov–Sachs (K–S) relationship:

$$(110)_{\text{bcc}} // (111)_{\text{fcc}}, [\bar{1}\bar{1}1]_{\text{bcc}} // [0\bar{1}1]_{\text{fcc}}$$

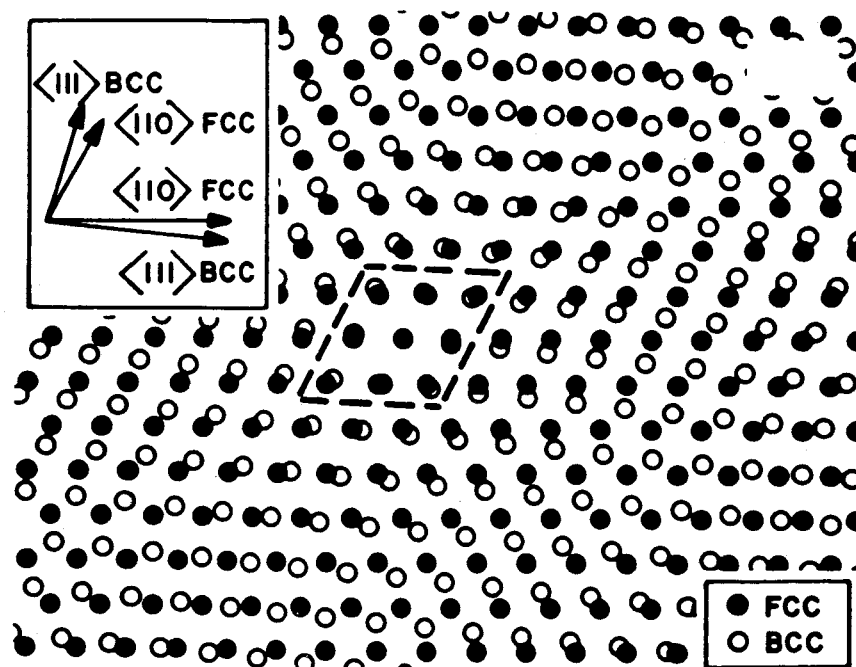


Fig. 3.38 Atomic matching across a $(111)_{\text{fcc}}/(110)_{\text{bcc}}$ interface bearing the NW orientation relationship for lattice parameters closely corresponding to the case of fcc and bcc iron (M.G. Hall *et al.*, *Surface Science*, 31 (1972) 257).

(The only difference between these two is a rotation in the closest-packed planes of 5.26° .) Figure 3.38 shows that the matching between a $\{111\}_{\text{fcc}}$ and $\{110\}_{\text{bcc}}$ plane bearing the N–W relationship is very poor. Good fit is restricted to small diamond-shaped areas that only contain $\sim 8\%$ of the interfacial atoms. A similar situation can be shown to exist for the K–S orientation relationship. Thus it can be seen that a coherent or semicoherent interface between the two phases is impossible for large interfaces parallel to $\{111\}_{\text{fcc}}$ and $\{110\}_{\text{bcc}}$. Such interfaces would be incoherent.

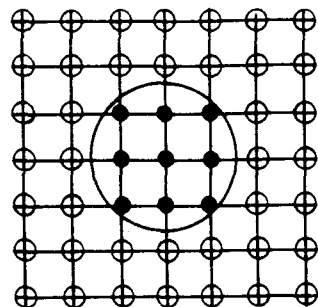
The degree of coherency can, however, be greatly increased if a macroscopically irrational interface is formed, (i.e. the indices of the interfacial plane in either crystal structure are not small integers). The detailed structure of such interfaces is, however, uncertain due to their complex nature^{8,9}.

3.4.2 Second-Phase Shape: Interfacial Energy Effects

In a two-phase microstructure one of the phases is often dispersed within the other, for example β -precipitates in an α -matrix. Consider for simplicity a system containing one β -precipitate embedded in a single α crystal, and assume for the moment that both the precipitate and matrix are strain free. Such a system will have a minimum free energy when the shape of the precipitate and its orientation relationship to the matrix are optimized to give the lowest total interfacial free energy ($\Sigma A_i \gamma_i$). Let us see how this can be achieved for different types of precipitate.

Fully Coherent Precipitates

If the precipitate (β) has the same crystal structure and a similar lattice parameter to the parent α phase the two phases can form low-energy coherent interfaces on all sides—provided the two lattices are in a parallel orientation relationship—as shown in Fig. 3.39a. This situation arises during the early stages of many precipitation hardening heat treatments, and the β phase is then termed a *fully coherent precipitate* or a *GP zone*. (GP for Guinier and Preston, who first discovered their existence. This discovery was made independently by Preston in the USA and Guinier in France, both employing X-ray diffraction techniques. Their work was later confirmed by transmission electron microscopy.) Since the two crystal structures match more or less perfectly across all interfacial planes the zone can be any shape and remain fully coherent. Thus a γ -plot of the α/β interfacial energy would be largely spherical and, ignoring coherency strains, the equilibrium shape of a zone should be a sphere. Figure 3.39b shows an example of GP zones, ~ 10 nm in diameter, in an Al–4 atomic % Ag alloy. The zones are a silver-rich fcc region within the aluminium-rich fcc matrix. Since the atomic diameters of aluminium and silver differ by only 0.7% the coherency strains make a negligible contribution to the total free energy of the alloy. In other systems such as Al–Cu where the atomic size difference is much larger strain energy is



(a)

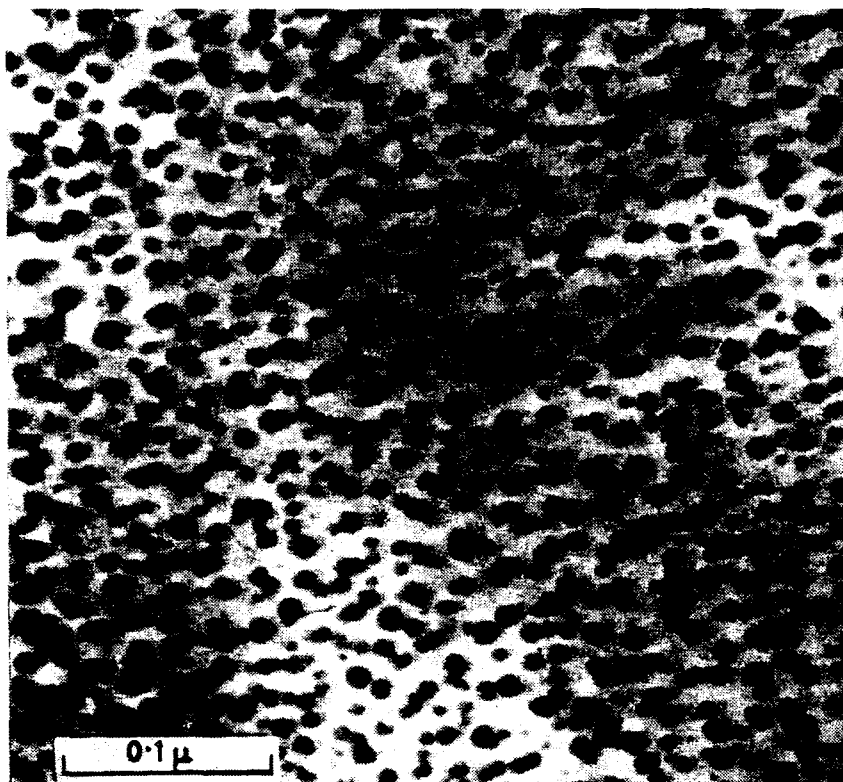


Fig. 3.39 (a) A zone with no misfit (○ Al, ● Ag, for example). (b) Electron micrograph of Ag-rich zones in an Al-4 atomic % Ag alloy ($\times 300\,000$). (After R.B. Nicholson, G. Thomas and J. Nutting, *Journal of the Institute of Metals*, 87 (1958–1959) 431.)

found to be more important than interfacial energy in determining the equilibrium shape of the zone. This point will be discussed further in Section 3.4.3.

Partially Coherent Precipitates

From an interfacial energy standpoint it is favourable for a precipitate to be surrounded by low-energy coherent interfaces. However, when the precipitate and matrix have different crystal structures it is usually difficult to find a lattice plane that is common to both phases. Nevertheless, for certain phase combinations there may be one plane that is more or less identical in each crystal, and by choosing the correct orientation relationship it is then possible for a low-energy coherent or semicoherent interface to be formed. There are, however, usually no other planes of good matching and the precipitate must consequently also be bounded by high-energy incoherent interfaces.

A γ -plot of the interfacial energy in this case could look like that in Fig. 3.40, i.e. roughly a sphere with two deep cusps normal to the coherent interface. The Wulff theorem would then predict the equilibrium shape to be a disc with a thickness/diameter ratio of γ_c/γ_i , where γ_c and γ_i are the energies of the (semi-) coherent and incoherent interfaces. Triangular, square, or hexagonal plate shapes would be predicted if the γ plot also contained smaller cusps at symmetrically disposed positions in the plane of the plate.

The precipitate shapes observed in practice may deviate from this shape for two main reasons. Firstly the above construction only predicts the equilibrium shape if misfit strain energy effects can be ignored. Secondly the precipitate may not be able to achieve an equilibrium shape due to constraints on how it can grow. For example disc-shaped precipitates may be much wider than the equilibrium shape if the incoherent edges grow faster than the broad faces.

Plate-like precipitates occur in many systems. For example the hcp γ' -phase in aged Al-4 atomic % Ag alloys forms as plates with semicoherent

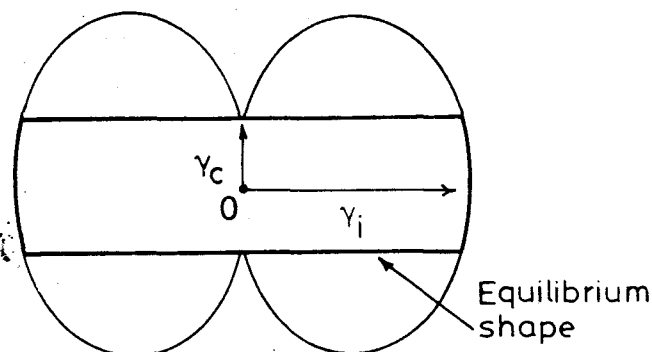


Fig. 3.40 A section through a γ -plot for a precipitate showing one coherent or semicoherent interface, together with the equilibrium shape (a disc).

broad faces parallel to the $\{111\}_\alpha$ matrix planes and with the usual hcp/fcc orientation relationship. The tetragonal θ' phase in aged Al-4 wt% Cu alloys are also plate-shaped, but in this case the broad faces of the plate (known as the *habit plane*) are parallel to $\{100\}_\alpha$ matrix planes. Figure 3.41 shows that the $\{100\}_\alpha$ planes are almost identical to the $(001)_{\theta'}$ plane so that the orientation relationship between the θ' and the aluminium-rich matrix (α) is

$$\begin{array}{l} (001)_{\theta'} // (001)_\alpha \\ [100]_{\theta'} // [100]_\alpha \end{array}$$

Examples of the precipitate shapes that are formed in these two systems are shown in Figs. 3.42 and 3.43. Note that as a result of the cubic symmetry of the aluminium-rich matrix there are many possible orientations for the precipitate plates within any given grain. This leads to a very characteristic crystallographic microstructure known after its discoverer as a *Widmanstätten morphology*.

Besides plate-like habits precipitates have also been observed to be lath-shaped (a plate elongated in one direction) and needle-like. For example the S phase in Al-Cu-Mg alloys forms as laths and the β' phase in Al-Mg-Si alloys as needles¹⁰. In both cases the precipitates are also crystallographically related to the matrix and produce a Widmanstätten structure.

Incoherent Precipitates

When the two phases have completely different crystal structures, or when the two lattices are in a random orientation, it is unlikely that any coherent or semicoherent interfaces form and the precipitate is said to be incoherent. Since the interfacial energy should be high for all interfacial planes, the γ -plot and the equilibrium inclusion shape will be roughly spherical. It is possible

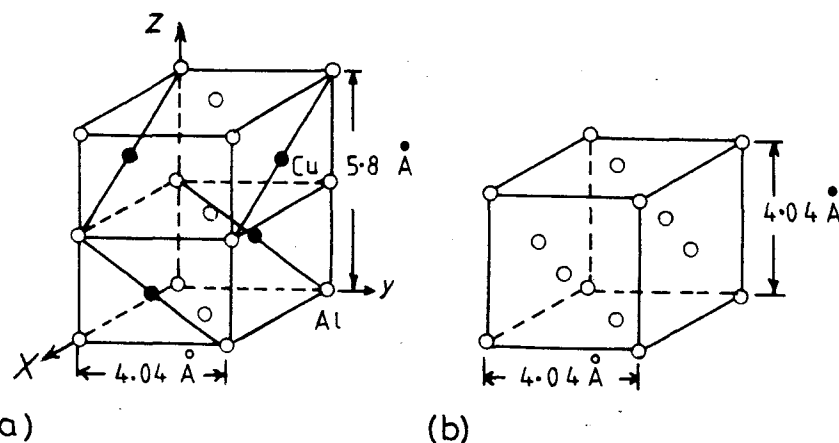


Fig. 3.41 (a) The unit cell of the θ' precipitate in Al-Cu alloys. (b) The unit cell of the matrix. (After J.M. Silcock, T.J. Heal and H.K. Hardy, *Journal of the Institute of Metals*, 82 (1953-1954) 239.)

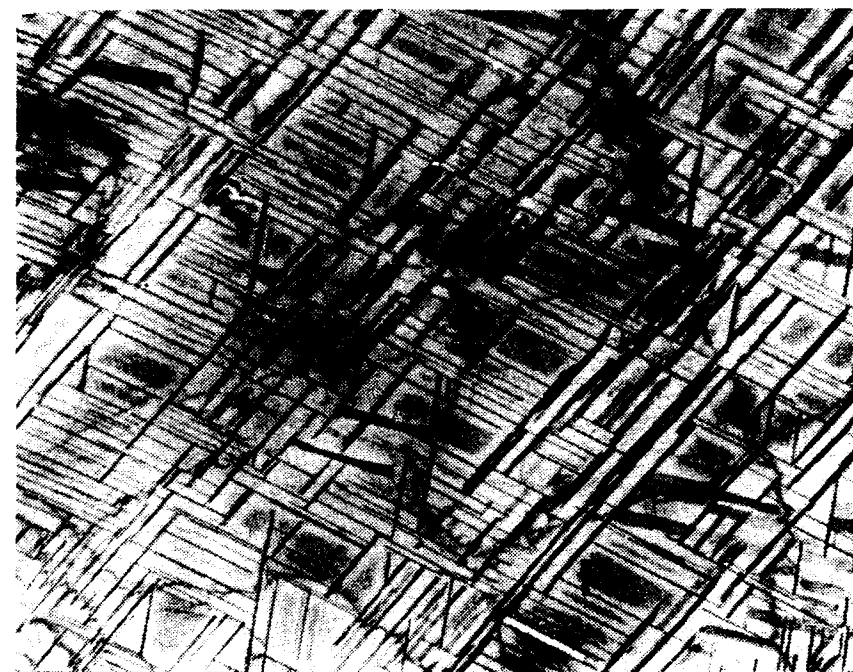


Fig. 3.42 Electron micrograph showing the Widmanstätten morphology of γ' precipitates in an Al-4 atomic % Ag alloy. GP zones can be seen between the γ' , e.g. at H ($\times 7000$). (R.B. Nicholson and J. Nutting, *Acta Metallurgica*, 9 (1961) 332.)

that certain crystallographic planes of the inclusion lie at cusps in the γ -plot so that polyhedral shapes are also possible. Such faceting, however, need not imply the existence of coherent or semicoherent interfaces.

The $\theta(\text{CuAl}_2)$ precipitate in Al-Cu alloys is an example of an incoherent precipitate, Fig. 3.44. It is found that there is an orientation relationship between the θ and aluminium matrix but this is probably because θ forms from the θ' phase and does not imply that θ is semicoherent with the matrix.

Precipitates on Grain Boundaries

Rather special situations arise when a second-phase particle is located on a grain boundary as it is necessary to consider the formation of interfaces with two differently oriented grains. Three possibilities now arise (Fig. 3.45): the precipitate can have (i) incoherent interfaces with both grains, (ii) a coherent or semicoherent interface with one grain and an incoherent interface with the other, or (iii) it can have a coherent or semicoherent interface with both grains. The first two cases are commonly encountered but the third possibility is unlikely since the very restrictive crystallographic conditions imposed by coherency with one grain are unlikely to yield a favourable orientation relationship towards the other grain.

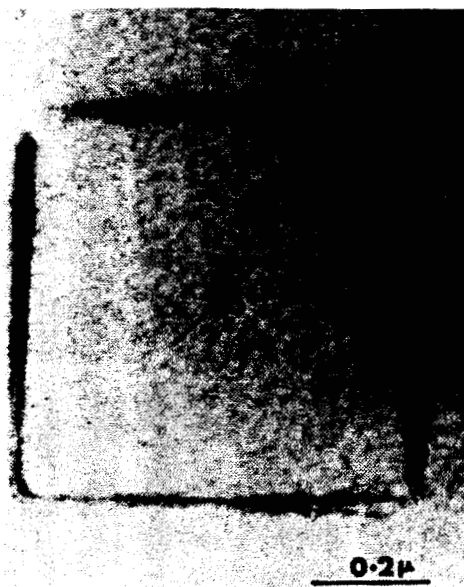


Fig. 3.43 Electron micrograph of a single coherent θ' plate in an Al-3.9 wt% Cu alloy aged 24 h at 200 °C. ($\times 80\,000$) (R. Sankaran and C. Laird, *Acta Metallurgica* 22 (1974) 957.)

The minimization of interfacial energy in these cases also leads to planar semicoherent (or coherent) interfaces and smoothly curved incoherent interfaces as before, but now the interfacial tensions and torques must also balance at the intersection between the precipitate and the boundary. (The shape that produces the minimum free energy can in fact be obtained by superimposing the γ plots for both grains in a certain way¹¹.) An example of a grain-boundary precipitate is shown in Fig. 3.46.

3.4.3 Second-Phase Shape: Misfit Strain Effects

Fully Coherent Precipitates

It was pointed out in the previous section that the equilibrium shape of a coherent precipitate or zone can only be predicted from the γ -plot when the misfit between the precipitate and matrix is small. When misfit is present the formation of coherent interfaces raises the free energy of the system on account of the elastic strain fields that arise. If this elastic strain energy is denoted by ΔG_s the condition for equilibrium becomes

$$\sum A_i \gamma_i + \Delta G_s = \text{minimum} \quad (3.35)$$



Fig. 3.44 Electron micrograph showing incoherent particles of θ in an Al-Cu alloy. (After G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972, from C. Laird.)

The origin of the coherency strains for a misfitting precipitate is demonstrated in Fig. 3.47. If the volume of matrix encircled in Fig. 3.47a is cut out and the atoms are replaced by smaller atoms the cut-out volume will undergo a uniform negative dilatational strain to an inclusion with a smaller lattice parameter, Fig. 3.47b. In order to produce a fully coherent precipitate the matrix and inclusion must be strained by equal and opposite forces as shown in Fig. 3.47c¹².

If the lattice parameters of the unstrained precipitate and matrix are a_p and a_m respectively the *unconstrained misfit* δ is defined by

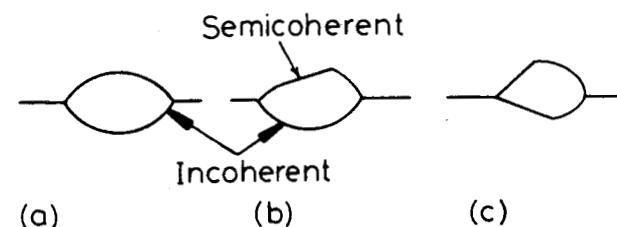


Fig. 3.45 Possible morphologies for grain boundary precipitates. Incoherent interfaces smoothly curved. Coherent or semicoherent interfaces planar.

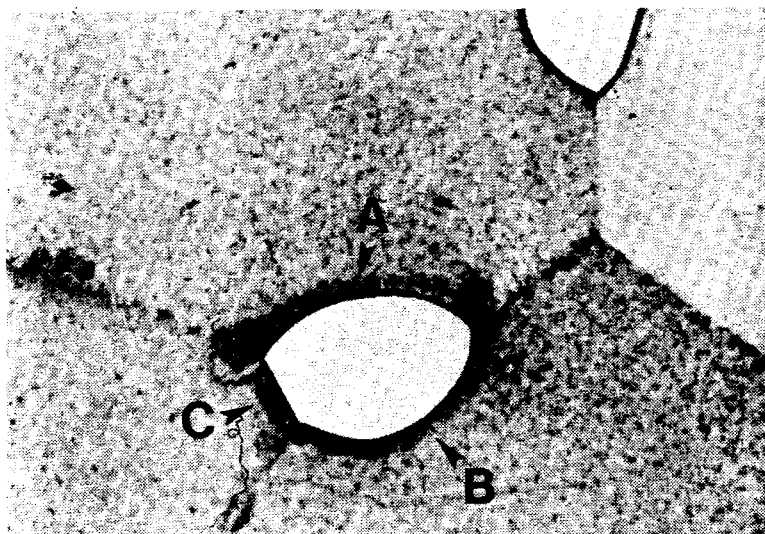


Fig. 3.46 An α precipitate at a grain boundary triple point in an $\alpha - \beta$ Cu-In alloy. Interfaces A and B are incoherent while C is semicoherent ($\times 310$). (After G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972.)

$$\delta = \frac{a_\beta - a_\alpha}{a_\alpha} \quad (3.36)$$

However, the stresses maintaining coherency at the interfaces distort the precipitate lattice, and in the case of a spherical inclusion the distortion is purely hydrostatic, i.e. it is uniform in all directions, giving a new lattice parameter a'_β . The *in situ* or *constrained misfit* ϵ is defined by

$$\epsilon = \frac{a'_\beta - a_\alpha}{a_\alpha} \quad (3.37)$$

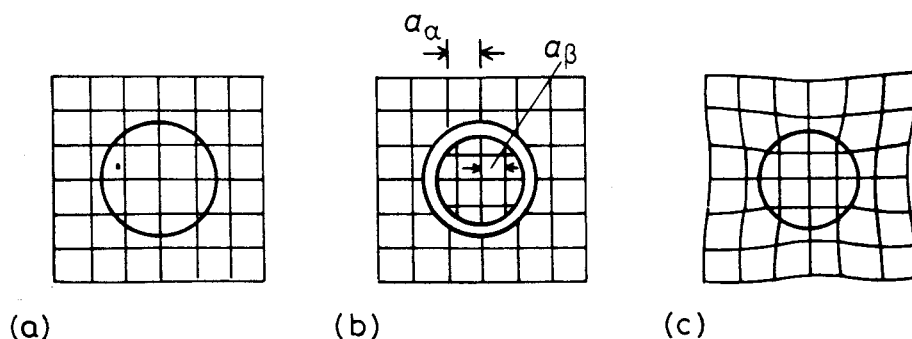


Fig. 3.47 The origin of coherency strains. The number of lattice points in the hole is conserved.

If the elastic moduli of the matrix and inclusion are equal and Poisson's ratio is $1/3$, ϵ and δ are simply related by

$$\epsilon = \frac{2}{3}\delta \quad (3.38)$$

In practice the inclusion has different elastic constants to the matrix, nevertheless ϵ still usually lies in the range $0.5\delta < \epsilon < \delta$.

When the precipitate is a thin disc the *in situ* misfit is no longer equal in all directions, but instead it is large perpendicular to the disc and almost zero in the plane of the broad faces, as shown in Fig. 3.48.

In general the total elastic energy depends on the shape and elastic properties of both matrix and inclusion. However, if the matrix is elastically isotropic and both precipitate and matrix have equal elastic moduli, the total elastic strain energy ΔG_s is independent of the shape of the precipitate, and assuming Poisson's ratio (ν) = $1/3$ it is given by

$$\Delta G_s \approx 4\mu\delta^2 \cdot V \quad (3.39)$$

where μ is the shear modulus of the matrix and V is the volume of the unconstrained hole in the matrix. Therefore coherency strains produce an elastic strain energy which is proportional to the volume of the precipitate and increases as the square of the lattice misfit (δ^2). If the precipitate and inclusion have different elastic moduli the elastic strain energy is no longer shape-independent but is a minimum for a sphere if the inclusion is hard and a disc if the inclusion is soft.

The above comments applied to isotropic matrices. In general, however, most metals are elastically anisotropic. For example, most cubic metals

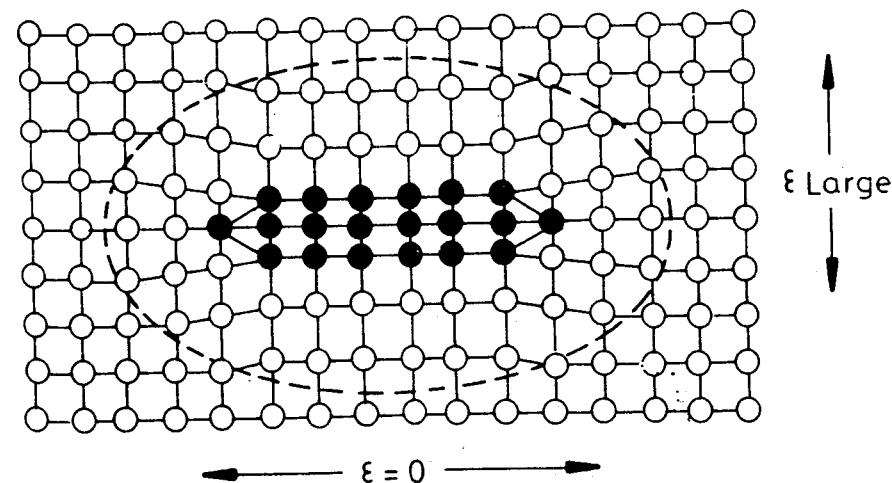


Fig. 3.48 For a coherent thin disc there is little misfit parallel to the plane of the disc. Maximum misfit is perpendicular to the disc.

(except molybdenum) are soft in $\langle 100 \rangle$ directions and hard in $\langle 111 \rangle$. The shape with a minimum strain energy under these conditions is a disc parallel to $\{100\}$ since most of the misfit is then accommodated in the soft directions perpendicular to the disc.

The influence of strain energy on the equilibrium shape of coherent precipitates can be illustrated by reference to zones in various aluminium-rich precipitation hardening alloys: Al-Ag, Al-Zn and Al-Cu. In each case zones containing 50–100% solute can be produced. Assuming the zone is pure solute the misfit can be calculated directly from the atomic radii as shown below.

Atom radius (Å)	Al:1.43	Ag:1.44	Zn:1.38	Cu:1.28
Zone misfit (δ)	—	+0.7%	-3.5%	-10.5%
Zone shape	—	sphere	sphere	disc

When $\delta < 5\%$ strain energy effects are less important than interfacial energy effects and spherical zones minimize the total free energy. For $\delta \geq 5\%$, as in the case of zones in Al-Cu, the small increase in interfacial energy caused by choosing a disc shape is more than compensated by the reduction in coherency strain energy.

Incoherent Inclusions

When the inclusion is incoherent with the matrix, there is no attempt at matching the two lattices and lattice sites are not conserved across the interface. Under these circumstances there are no coherency strains. Misfit strains can, however, still arise if the inclusion is the wrong size for the hole it is located in, Fig. 3.49. In this case the lattice misfit δ has no significance and it is better to consider the volume misfit Δ as defined by

$$\Delta = \frac{\Delta V}{V} \quad (3.40)$$

where V is the volume of the unconstrained hole in the matrix and $(V - \Delta V)$ the volume of the unconstrained inclusion. (For a coherent spherical inclusion

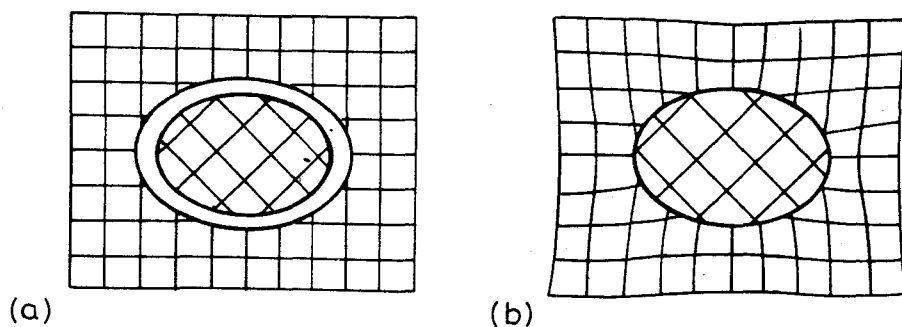


Fig. 3.49 The origin of misfit strain for an incoherent inclusion (no lattice matching).

the volume misfit and the linear lattice misfit are related by $\Delta = 3\delta$. But for a non-coherent sphere the number of lattice sites within the hole is not preserved (see Fig. 3.49) and in this case $\Delta \neq 3\delta$.) When the matrix hole and inclusion are constrained to occupy the same volume the elastic strain fields again result as shown in Fig. 3.49b. The elasticity problem in this case has been solved for spheroidal inclusions which are described by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{c^2} = 1 \quad (3.41)$$

Nabarro¹³ gives the elastic strain energy for a homogeneous incompressible inclusion in an isotropic matrix as

$$\Delta G_s = \frac{2}{3} \mu \Delta^2 \cdot V \cdot f(c/a) \quad (3.42)$$

where μ is the shear modulus of the matrix. Thus the elastic strain energy is proportional to the square of the volume misfit Δ^2 . The function $f(c/a)$ is a factor that takes into account the shape effects and is shown in Fig. 3.50. Notice that, for a given volume, a sphere ($c/a = 1$) has the highest strain energy while a thin, oblate spheroid ($c/a \rightarrow 0$) has a very low strain energy, and a needle shape ($c/a = \infty$) lies between the two. If elastic anisotropy is included¹⁴ it is found that the same general form for $f(c/a)$ is preserved and only small changes in the exact values are required. Therefore the equilibrium shape of an incoherent inclusion will be an oblate spheroid with c/a value that balances the opposing effects of interfacial energy and strain

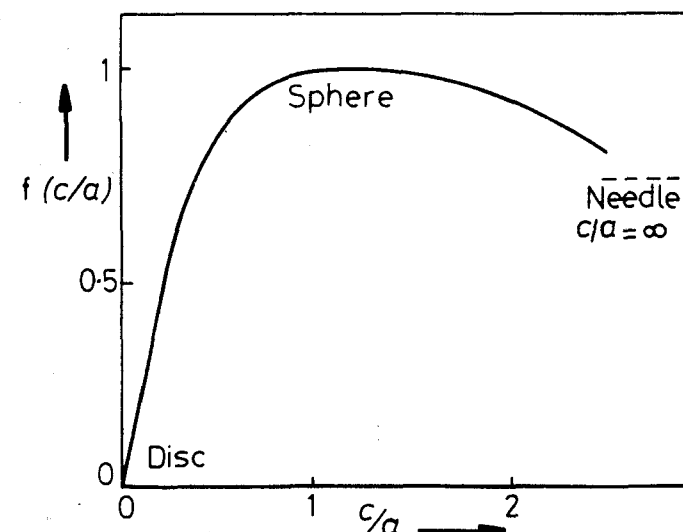


Fig. 3.50 The variation of misfit strain energy with ellipsoid shape, $f(c/a)$. (After F.R.N. Nabarro, *Proceedings of the Royal Society A*, 175 (1940) 519.)

energy. When Δ is small interfacial energy effects should dominate and the inclusion should be roughly spherical.

Plate-Like Precipitates

Consider a plate-like precipitate with coherent broad faces and incoherent or semicoherent edges. Fig. 3.51. (The criterion for whether these interfaces are coherent or semicoherent is discussed in the following section.) Misfit across the broad faces then results in large coherency strains parallel to the plate, but no coherency strains will exist across the edges. The *in situ* misfit across the broad faces increases with increasing plate thickness which leads to greater strains in the matrix and higher shear stresses at the corners of the plates¹⁵. Eventually it becomes energetically favourable for the broad faces to become semicoherent. Thereafter the precipitate behaves as an incoherent inclusion with comparatively little misfit strain energy. An example of a precipitate that can be either coherent or semicoherent in this way is θ' in Al-Cu alloys (see Section 5.5.1).

3.4.4 Coherency Loss

Precipitates with coherent interfaces have a low interfacial energy, but in the presence of misfit, they are associated with a coherency strain energy. On the other hand, if the same precipitate has non-coherent interfaces it will have a higher interfacial energy but the coherency strain energy will be absent. Let us now consider which state produces the lowest total energy for a spherical precipitate with a misfit δ and a radius r .

The free energy of a crystal containing a fully coherent spherical precipitate has contributions from (i) the coherency strain energy given by Equation 3.39, and (ii) the chemical interfacial energy γ_{ch} . The sum of these two terms is given by

$$\Delta G(\text{coherent}) = 4\mu\delta^2 \cdot \frac{4}{3}\pi r^3 + 4\pi r^2 \cdot \gamma_{ch} \quad (3.43)$$

If the same precipitate has incoherent or semicoherent interfaces that completely relieve the unconstrained misfit there will be no misfit energy, but

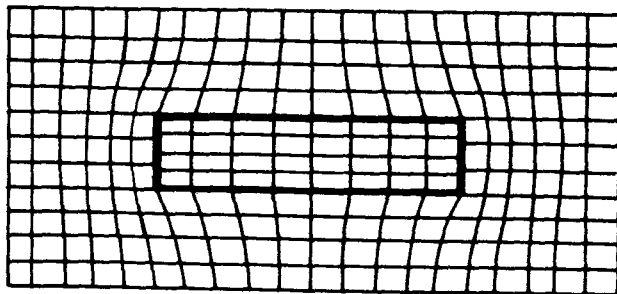


Fig. 3.51 Coherency strains caused by the coherent broad faces of θ' precipitates.

there will be an extra structural contribution to the interfacial energy γ_{st} . The total energy in this case is given by

$$\Delta G(\text{non-coherent}) = 0 + 4\pi r^2(\gamma_{ch} + \gamma_{st}) \quad (3.44)$$

For a given δ , ΔG (coherent) and ΔG (non-coherent) vary with r as shown in Fig. 3.52. When small, therefore, the coherent state gives the lowest total energy, while it is more favourable for large precipitates to be semicoherent or incoherent (depending on the magnitude of δ). At the critical radius (r_{crit}) $\Delta G(\text{coherent}) = \Delta G(\text{non-coherent})$ giving

$$r_{crit} = \frac{3\gamma_{st}}{4\mu\delta^2} \quad (3.45)$$

If we assume that δ is small, a semicoherent interface will be formed with a structural energy $\gamma_{st} \propto \delta$. In which case

$$r_{crit} \propto \frac{1}{\delta} \quad (3.46)$$

If a coherent precipitate grows, during ageing for example, it should lose coherency when it exceeds r_{crit} . However, as shown in Fig. 3.53 loss of coherency requires the introduction of dislocation loops around the precipitate and in practice this can be rather difficult to achieve. Consequently coherent precipitates are often found with sizes much larger than r_{crit} .

There are several ways in which coherency may be lost and some of them are illustrated in Fig. 3.54. The most straightforward way is for a dislocation loop to be punched out at the interface as shown in Fig. 3.54a. This requires the stresses at the interface to exceed the theoretical strength of the matrix. However, it can be shown that the punching stress p_s is independent of the precipitate size and depends only on the constrained misfit ϵ . If the shear modulus of the matrix is μ

$$p_s = 3\mu\epsilon \quad (3.47)$$

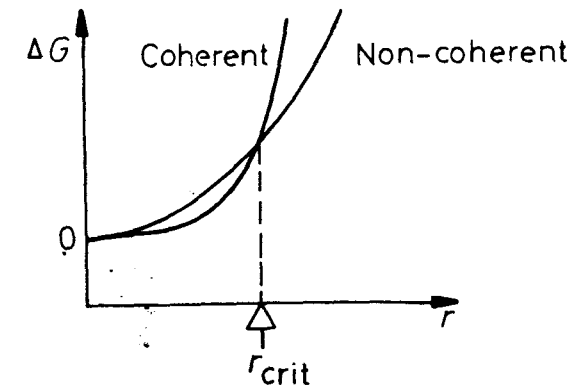


Fig. 3.52 The total energy of matrix + precipitate v. precipitate radius for spherical coherent and non-coherent (semicoherent or incoherent) precipitates.

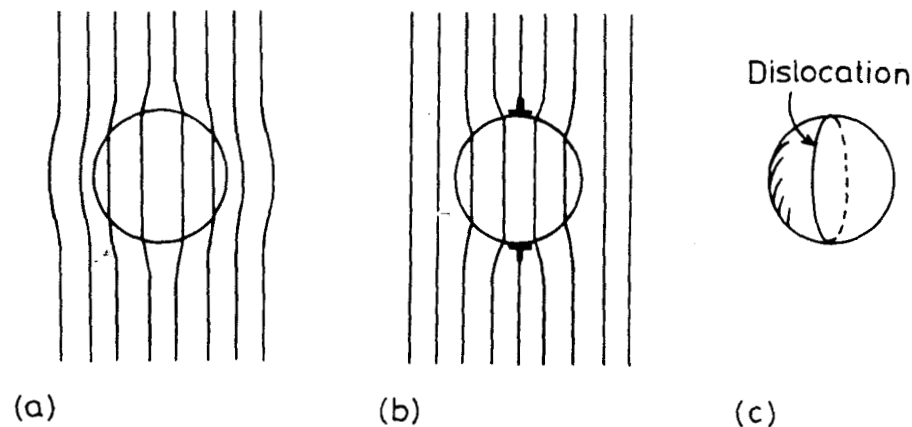


Fig. 3.53 Coherency loss for a spherical precipitate. (a) Coherent. (b) Coherency strains replaced by dislocation loop. (c) In perspective.

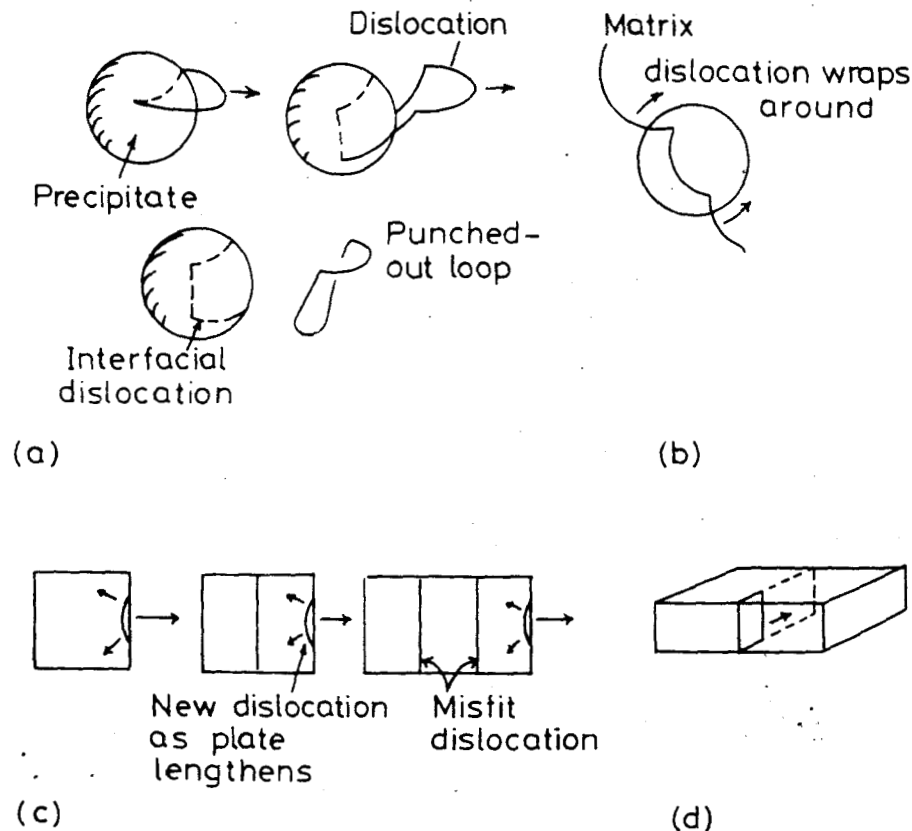


Fig. 3.54 Mechanisms for coherency loss. (a) Dislocation punching from interface. (b) Capture of matrix dislocation. (c) Nucleation at edge of plate repeated as plate lengthens. (d) Loop expansion by vacancy condensation in the precipitate.

It has been estimated that the critical value of ϵ that can cause the theoretical strength of the matrix to be exceeded is approximately given by

$$\epsilon_{\text{crit}} = 0.05 \quad (3.48)$$

Consequently precipitates with a smaller value of ϵ cannot lose coherency by this mechanism, no matter how large.

There are several alternative mechanisms but all require the precipitate to reach a larger size than r_{crit} ¹⁶. For example, the precipitate can attract a matrix dislocation with a suitable Burgers vector, and cause it to wrap itself around the precipitate, Fig. 3.54b. This mechanism is difficult in annealed specimens but is assisted by mechanical deformation.

In the case of plate-like precipitates the situation is different and it is now possible for the high stresses at the edges of the plates to nucleate dislocations by exceeding the theoretical strength of the matrix. The process can be repeated as the plate lengthens so as to maintain a roughly constant interdislocation spacing, Fig. 3.54c. Another mechanism that has been observed for plate-like precipitates is the nucleation of dislocation loops within the precipitate¹⁷. Vacancies can be attracted to coherent interfaces¹⁸ and 'condense' to form a prismatic dislocation loop which can expand across the precipitate, as shown in Fig. 3.54d.

3.4.5 Glissile Interfaces

In the treatment of semicoherent interfaces that has been presented in the previous sections it has been assumed that the misfit dislocations have Burgers vectors parallel to the interfacial plane. This type of interface is referred to as epitaxial. Glide of the interfacial dislocations cannot cause the interface to advance and the interface is therefore non-glissile. It is however possible, under certain circumstances, to have glissile semicoherent interfaces which can advance by the coordinated glide of the interfacial dislocations. This is possible if the dislocations have a Burgers vector that can glide on matching planes in the adjacent lattices as illustrated in Fig. 3.55. The slip planes must be continuous across the interface, but not necessarily parallel. Any gliding dislocation shears the lattice above the slip plane relative to that below by the Burgers vector of the dislocation. In the same way the gliding of the dislocations in a glissile interface causes the receding lattice, α say, to be *sheared* into the β -structure.

As an aid to understanding the nature of glissile boundaries consider two simple cases. The first is the low-angle symmetric tilt boundary, shown in Figs. 3.7a and 3.11. In this case the Burgers vectors are all pure edge in nature and as they glide one grain is rotated into the other grain. Strictly speaking this is not an interphase interface as there is no change in crystal structure, just a rotation of the lattice. A slightly more complex example of a glissile interface between two different lattices is that which can arise between

Interfacial dislocations

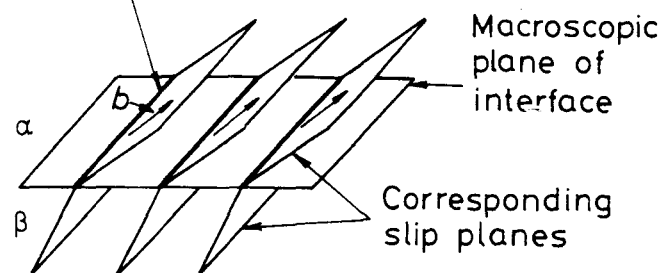


Fig. 3.55 The nature of a glissile interface.

the cubic and hexagonal close-packed lattices. To understand the structure of this interface requires a slight digression to consider the nature of *Shockley partial dislocations*.

Both fcc and hcp lattices can be formed by stacking close-packed layers of atoms one above the other. If the centres of the atoms in the first layer are denoted as A-positions, the second layer of atoms can be formed either by filling the B-positions, or C-positions as shown in Fig. 3.56. Either position produces the same atomic configuration at this stage. Let us assume therefore that the atoms in the second layer occupy B-sites. There are now two non-equivalent ways of stacking the third layer. If the third layer is placed directly above the first layer the resulting stacking sequence is ABA and the addition of further layers in the same sequence ABABABABAB... has hexagonal symmetry and is known as a hexagonal close-packed arrangement. The unit cell and stacking sequence of this structure are shown in Fig. 3.57. The close-packed plane can therefore be indexed as (0001) and the close-packed directions are of the type $\langle 11\bar{2}0 \rangle$.

If the atoms in the third layer are placed on the C-sites to form ABC and the same sequence is then repeated, the stacking sequence becomes ABCABCAB... which produces a cubic close-packed arrangement with a face-centred cubic unit cell as shown in Fig. 3.58. The close-packed atomic planes in this case become the $\{111\}$ type and the close-packed directions the $\langle 110 \rangle$ type.

In terms of the fcc unit cell the distance between the B- and C-sites measured parallel to the close-packed planes corresponds to vectors of the type $\frac{a}{2} \langle 11\bar{2} \rangle$. Therefore if a dislocation with a Burgers vector $\frac{a}{2} [11\bar{2}]$ glides between two (111) layers of an fcc lattice, say layers 4 and 5 in Fig. 3.59, all layers above the glide plane (5, 6, 7...) will be shifted relative to those below the glide plane by a vector $\frac{a}{2} [11\bar{2}]$. Therefore all atoms above the glide plane in B-sites are moved to C-sites, atoms in C-sites move to A-sites, and atoms in A-sites move to B-sites, as shown in Fig. 3.59. This type of dislocation with $b = \frac{a}{2} \langle 11\bar{2} \rangle$ is known as Shockley *partial* dislocation. They are called partial dislocations because vectors of the type $\frac{a}{2} \langle 11\bar{2} \rangle$ do not connect lattice

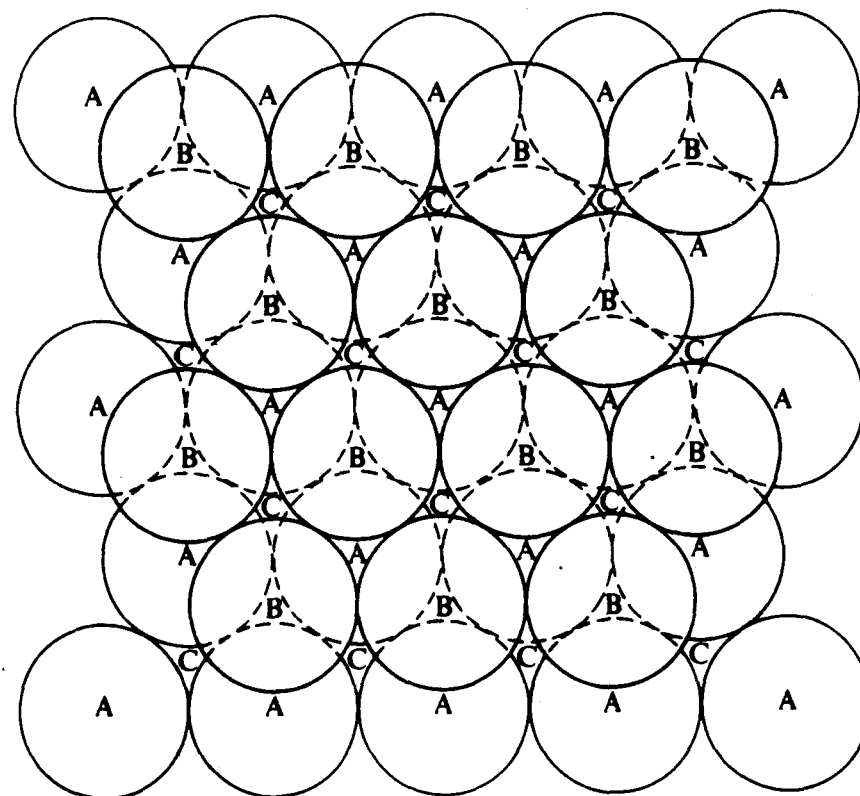


Fig. 3.56 The location of A, B and C sites in a close-packed layer of atoms. See also Figs. 3.57 and 3.58. (After J.W. Martin and R.D. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, Cambridge, 1976.)

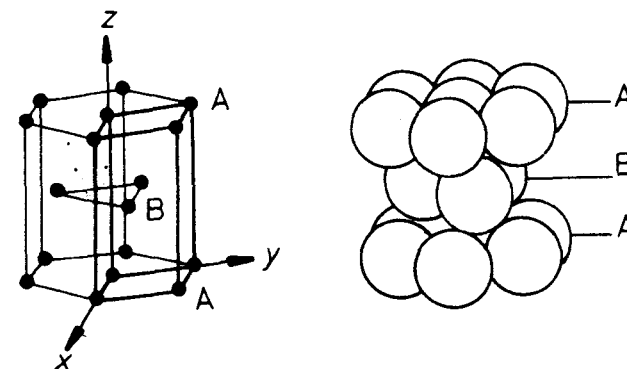


Fig. 3.57 A hexagonal close-packed unit cell and stacking sequence.

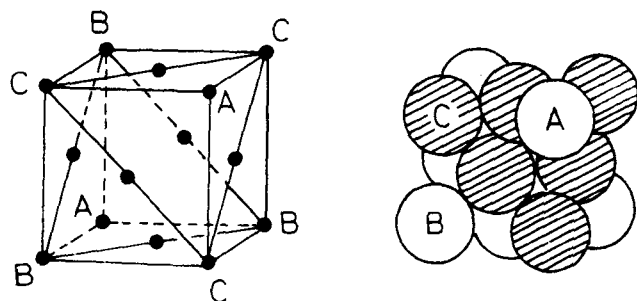
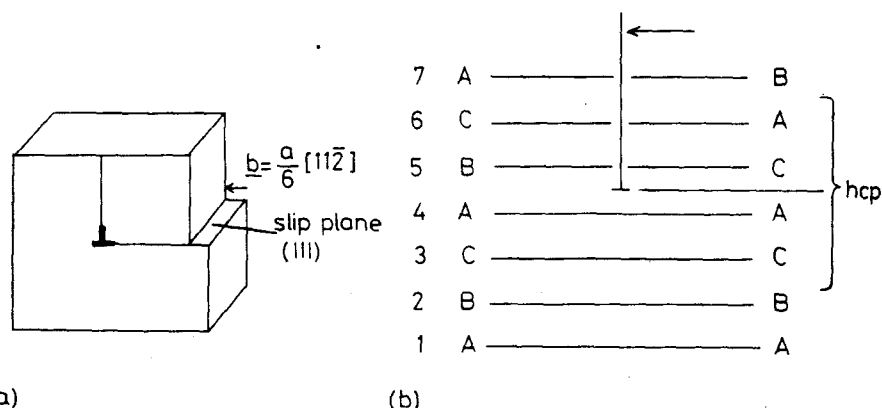


Fig. 3.58 A cubic close-packed structure showing fcc unit cell and stacking sequence.

Fig. 3.59 (a) An edge dislocation with a Burgers vector $\mathbf{b} = \frac{a}{6} [11\bar{2}]$ on (111). (Shockley partial dislocation.) (b) The same dislocation locally changes the stacking sequence from fcc to hcp.

points in the fcc structure. The gliding of Shockley partial dislocations therefore disrupts the crystal lattice and causes a stacking fault over the area of glide plane swept by the dislocation. Figure 3.59 shows that the nature of this fault is such that four layers of material are converted into a hexagonal close-packed sequence CACA. Therefore in thermodynamically stable fcc lattices the stacking fault is a region of high free energy. On the other hand if the fcc lattice is only metastable with respect to the hcp structure the stacking fault energy will be effectively negative and the gliding of Shockley partial dislocations will decrease the free energy of the system.

Consider now the effect of passing another $\frac{a}{6} [11\bar{2}]$ dislocation between layers 6 and 7 as shown in Fig. 3.60. It can be seen that the region of hcp stacking is now extended by a further two layers. Therefore a sequence of Shockley partial dislocations between every other (111) plane will create a glissile interface separating fcc and hcp crystals, Fig. 3.61.

The glide planes of the interfacial dislocations are continuous from the fcc to the hcp lattice and the Burgers vectors of the dislocations, which necessarily lie in the glide plane, are at an angle to the *macroscopic* interfacial

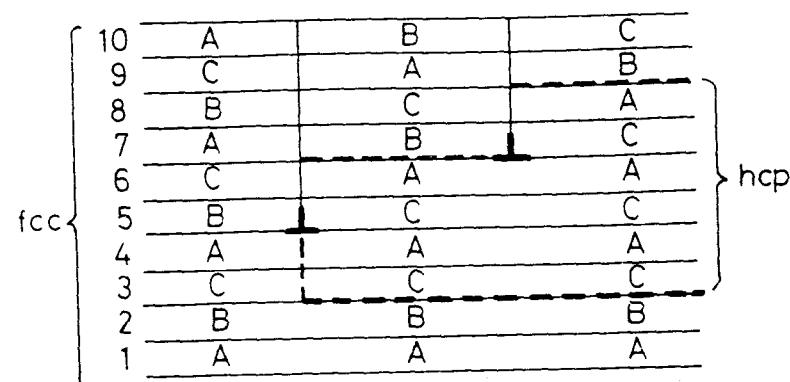


Fig. 3.60 Two Shockley partial dislocations on alternate (111) planes create six layers of hcp stacking.

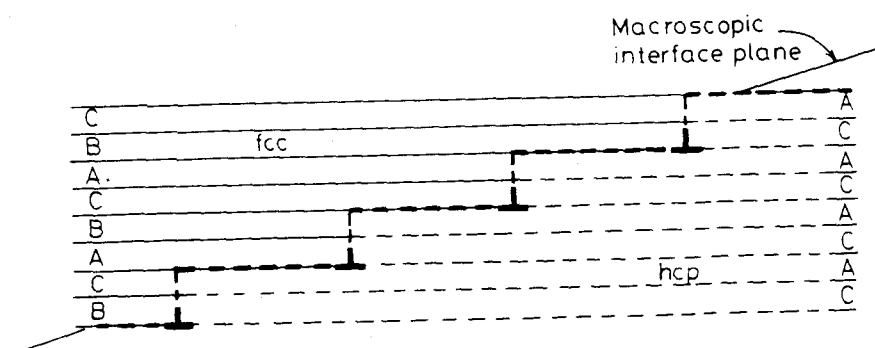


Fig. 3.61 An array of Shockley partial dislocations forming a glissile interface between fcc and hcp crystals.

plane. If the dislocation network glides into the fcc crystal it results in a transformation of $\text{fcc} \rightarrow \text{hcp}$, whereas a $\text{hcp} \rightarrow \text{fcc}$ transformation can be brought about by the reverse motion. Macroscopically the interfacial plane lies at an angle to the (111) or (0001) planes and need not be parallel to any low-index plane, i.e. it can be *irrational*. Microscopically, however, the interface is stepped into planar coherent facets parallel to $(111)_{\text{fcc}}$ and $(0001)_{\text{hcp}}$ with a step height the thickness of two closed-packed layers.

An important characteristic of glissile dislocation interfaces is that they can produce a macroscopic *shape change* in the crystal. This is illustrated for the $\text{fcc} \rightarrow \text{hcp}$ transformation in Fig. 3.62a. If a single fcc crystal is transformed into an hcp crystal by the passage of the same Shockley partial over every (111) plane then there is a macroscopic shape change, in this case a simple shear, as shown. There are, however, two other Shockley partials which can also be used to transform $\text{fcc} \rightarrow \text{hcp}$ stacking, and if the transformation is achieved using all three partials in equal numbers there will be no overall shape change, Fig. 3.62b.

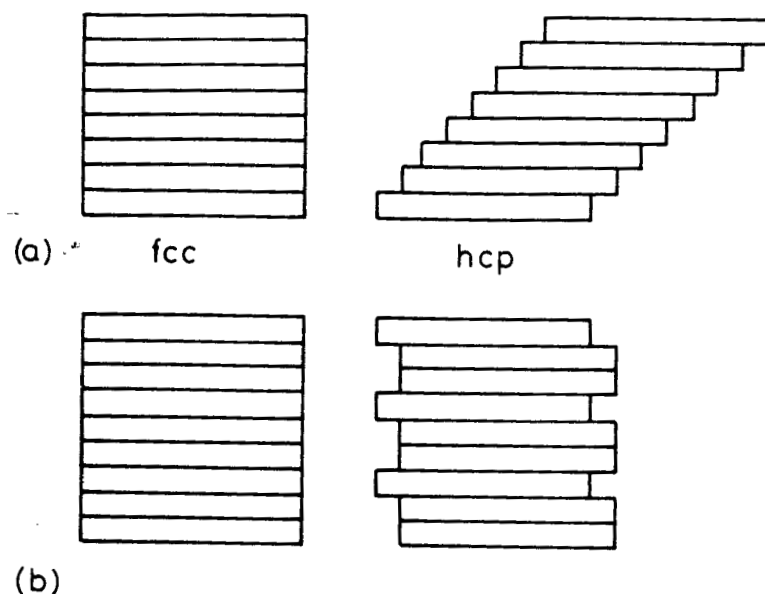


Fig. 3.62 Schematic representation of the different ways of shearing cubic close-packed planes into hexagonal close-packed (a) using only one Shockley partial, (b) using equal numbers of all three Shockley partials.

The formation of martensite in steel and other alloy systems occurs by the motion of glissile-dislocation interfaces. These transformations are characterized by a macroscopic shape change and no change in composition. Usually, however, the interface must be more complex than the fcc/hcp case discussed above, although the same principles will still apply. Martensitic transformations are dealt with further in Chapter 6.

3.4.6 Solid/Liquid Interfaces¹⁹

Many of the ideas that were discussed with regard to solid/vapour interfaces can be carried over to solid/liquid interfaces, only now the low density vapour phase is replaced by a high density liquid, and this has important consequences for the structure and energy of the interface.

There are basically two types of atomic structure for solid/liquid interfaces. One is essentially the same as the solid/vapour interfaces described in Section 3.1, i.e. an atomically flat close-packed interface, Fig. 3.63a. In this case the transition from liquid to solid occurs over a rather narrow transition zone approximately one atom layer thick. Such interfaces can also be described as smooth, faceted, or sharp. The other type is an atomically *diffuse* interface, Fig. 3.63b, in which the transition from liquid to solid occurs over several atom layers. Thus there is a gradual weakening of the interatomic bonds and an increasing disorder across the interface into the bulk liquid

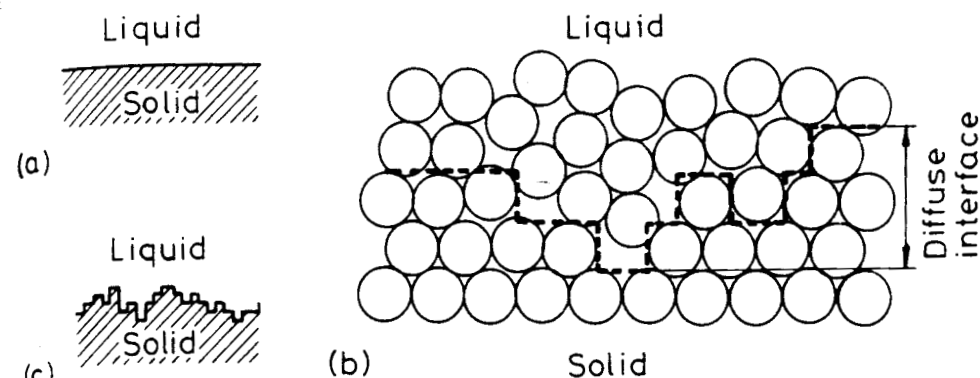


Fig. 3.63 Solid/liquid interfaces: (a) atomically smooth, (b) and (c) atomically rough, or diffuse interfaces. (After M.C. Flemings, *Solidification Processing*, McGraw-Hill, New York, 1974.)

phase; or in thermodynamic terms, enthalpy and entropy gradually change from bulk solid to bulk liquid values across the interface as shown in Fig. 3.64. When the solid and liquid are in equilibrium (at T_m) the high enthalpy of the liquid is balanced by a high entropy so that both phases have the same free energy. In the interface, however, the balance is disturbed thereby giving rise to an excess free energy, γ_{SL} .

Diffuse interfaces are also known as rough or non-faceted. The dotted line

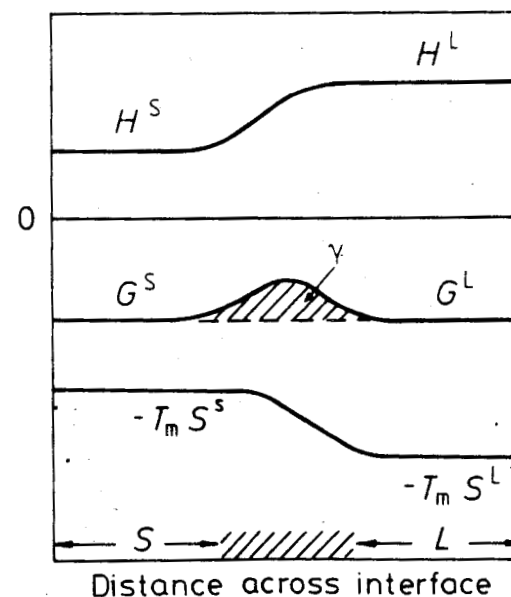


Fig. 3.64 The variation of H , $-T_m S$ and G across the solid/liquid interface at the equilibrium melting temperature T_m , showing the origin of the solid/liquid interfacial energy γ .

in Fig. 3.63b is an attempt to show the rough nature of the interface by dividing the atoms into the 'solid' and 'liquid'. If this is done the schematic representation of Fig. 3.63c can be used.

The type of structure chosen by a particular system will be that which minimizes the interfacial free energy. According to a simple theory developed by Jackson²⁰ the optimum atomic arrangement depends mainly on the latent heat of fusion (L_f) relative to the melting temperature (T_m). This theory predicts that there is a critical value of $L_f/T_m \approx 4R$ above which the interface should be flat and below which it should be diffuse. Most metals have $L_f/T_m \approx R$ and are therefore predicted to have rough interfaces. On the other hand some intermetallic compounds and elements such as Si, Ge, Sb as well as most non-metals have high values of L_f/T_m and generally have flat close-packed interfaces. If the model is applied to solid/vapour interfaces L_s (the heat of sublimation) should be used instead of L_f and then flat surfaces are predicted even for metals, in agreement with observations.

If the broken-bond model is applied to the calculation of the energy of a solid/liquid interface it can be argued that the atoms in the interface are roughly half bonded to the solid and half to the liquid so that the interfacial enthalpy should be $\sim 0.5 L_f/N_a$ per atom. This appears to compare rather favourably with experimentally measured values of γ_{SL} which are $\sim 0.45 L_f/N_a$ per atom for most metals. However the agreement is probably only fortuitous since entropy effects should also be taken into account, Fig. 3.64.

Some experimentally determined values of γ_{SL} are listed in Table 3.4.

Table 3.4 Experimentally Determined Solid/Liquid Interfacial Free Energies

Values selected from D. Turnbull, *Journal of Applied Physics*, Vol. 21: 1022(1950).

Material	T_m/K	$\gamma_{SL}/\text{mJ m}^{-2}$
Sn	505.7	54.5
Pb	600.7	33.3
Al	931.7	93
Ag	1233.7	126
Au	1336	132
Cu	1356	177
Mn	1493	206
Ni	1725	255
Co	1763	234
Fe	1803	204
Pd	1828	209
Pt	2043	240

These values were determined by indirect means from homogeneous nucleation experiments (see Chapter 4) and may contain systematic errors. Comparison of Tables 3.2 and 3.3 indicates $\gamma_{SL} \approx 0.30\gamma_b$ (for a grain boundary). More direct experiments²¹ imply that $\gamma_{SL} \approx 0.45\gamma_b (\approx 0.15\gamma_{SV})$. Another useful empirical relationship is that

$$\gamma_{SV} > \gamma_{SL} + \gamma_{LV}$$

which means that for a solid metal close to T_m it is energetically favourable for the surface to melt and replace the solid/vapour interface with solid/liquid and liquid/vapour interfaces.

It is found experimentally that the free energies of diffuse interfaces do not vary with crystallographic orientation, i.e. γ -plots are spherical²². Materials with atomically flat interfaces, however, show strong crystallographic effects and solidify with low-index close-packed facets, Fig. 3.65.

3.5 Interface Migration

The great majority of phase transformations in metals and alloys occur by a process known as nucleation and growth, i.e. the new phase (β) first appears

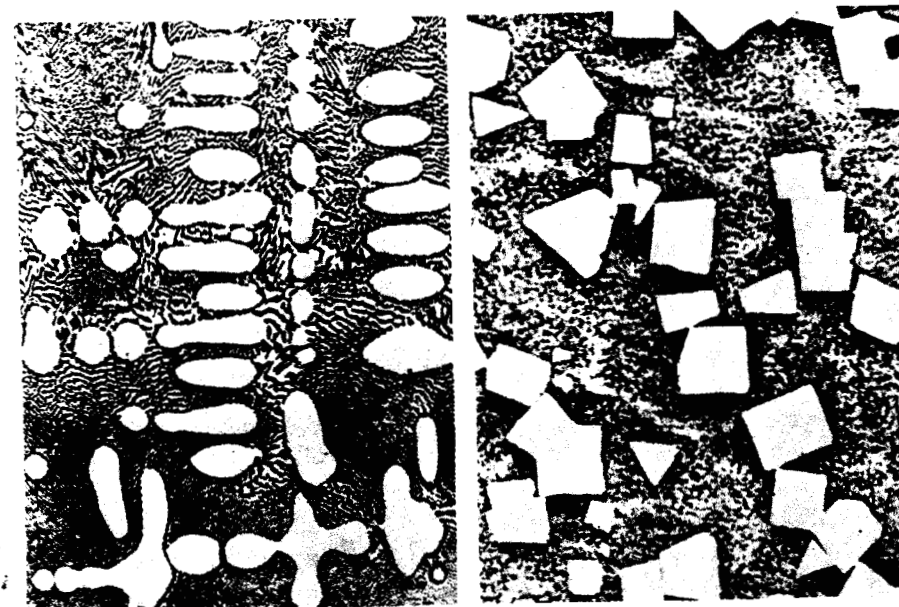


Fig. 3.65 Examples of solid-liquid interface structure in metallic systems. (a) Non-faceted dendrites of silver in a copper-silver eutectic matrix ($\times 330$); (b) faceted cuboids of β' -SnSb compound in a matrix of Sn-rich material ($\times 110$). (After G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972.)

at certain sites within the metastable parent (α) phase (nucleation) and this is subsequently followed by the growth of these nuclei into the surrounding matrix. In other words, an interface is created during the nucleation stage and then migrates into the surrounding parent phase during the growth stage. This type of transformation is therefore essentially heterogeneous, i.e. at any time during the transformation the system can be divided into parent and product phases. The nucleation stage is very important and determines many features of the transformation. However, most of the transformation product is formed during the growth stage by the transfer of atoms across the moving parent/product interface.

There are basically two different types of interface: glissile and non-glissile. Glissile interfaces migrate by dislocation glide that results in the shearing of the parent lattice into the product. The motion of glissile interfaces is relatively insensitive to temperature and is therefore known as *athermal* migration. Most interfaces are non-glissile and migrate by the more or less random jumps of individual atoms across the interface in a similar way to the migration of a random high-angle grain boundary. The extra energy that the atom needs to break free of one phase and attach itself to the other is supplied by *thermal activation*. The migration of non-glissile interfaces is therefore extremely sensitive to temperature.

A convenient way of classifying nucleation and growth transformations is to divide them according to the way in which the product grows. Therefore two major groupings can be made by dividing the transformations according to whether growth involves glissile or non-glissile interfaces. Transformations produced by the migration of a glissile interface are referred to as *military transformations*. This emphasizes the analogy between the coordinated motion of atoms crossing the interface and that of soldiers moving in ranks on the parade ground. In contrast the uncoordinated transfer of atoms across a non-glissile interface results in what is known as a *civilian transformation*.

During a military transformation the nearest neighbours of any atom are essentially unchanged. Therefore the parent and product phases must have the same composition and no diffusion is involved in the transformation. Martensitic transformations belong to this group. Glissile interfaces are also involved in the formation of mechanical twins and twinning therefore has much in common with martensitic transformations.

During civilian transformations the parent and product may or may not have the same composition. If there is no change in composition, e.g. the $\alpha \rightarrow \gamma$ transformation in pure iron, the new phase will be able to grow as fast as the atoms can cross the interface. Such transformations are said to be *interface controlled*. When the parent and product phases have different compositions, growth of the new phase will require long-range diffusion. For example, the growth of the B-rich β phase into the A-rich α phase shown in Fig. 3.66 can only occur if diffusion is able to transport A away from, and B towards the advancing interface. If the interfacial reaction is fast, i.e. the transfer of atoms across the interface is an easy process, the rate at which the

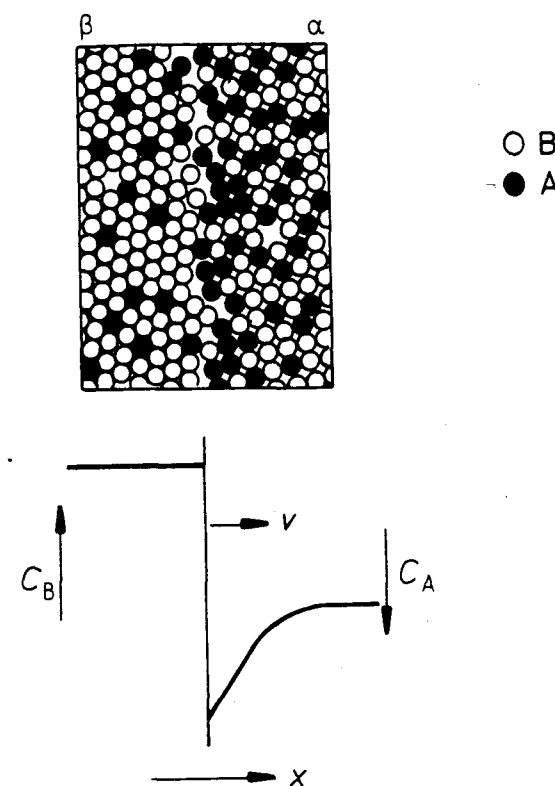


Fig. 3.66 Composition changes in a substitutional alloy caused by interface migration when the two adjoining phases have different compositions.

β phase can grow will be governed by the rate at which lattice diffusion can remove the excess atoms from ahead of the interface. This is therefore known as *diffusion-controlled* growth. However, if for some reason the interfacial reaction is slow, the growth rate will be governed by the interface kinetics. Under these circumstances growth is said to be *interface controlled* and a very small concentration gradient in the matrix is sufficient to provide the necessary flux of atoms to and from the interface. It is also possible that the interface reaction and diffusion process occur at similar rates in which case the interface is said to migrate under *mixed control*.

The above discussion of interface migration and classification of nucleation and growth transformations (also known as heterogeneous transformations) is summarized in Table 3.5, together with some examples of each class. This classification is adapted from that first proposed by Christian^{23,24}. Non-glissile interfaces can be considered to include solid/liquid and solid/vapour interfaces as well as solid/solid (coherent, semicoherent and incoherent) interfaces. Therefore solidification and melting can be included in the classification of civilian transformations under diffusion control (although the concept of diffusion may sometimes need to be extended to include the diffusion of

Table 3.5 Classification of Nucleation and Growth Transformations

Adapted from J.W. Christian, 'Phase transformations in metals and alloys—an introduction', in *Phase Transformations*, Vol. 1, p. 1, Institute of Metallurgists, 1979.

Type	Military	Civilian
Effect of temperature change	Athermal	Thermally activated
Interface type	Glissile (coherent or semicoherent)	Non-Glissile (coherent, semicoherent, incoherent, solid/liquid or solid/vapour)
Composition of parent and product phases	Same composition	Different compositions
Nature of diffusion processes	No diffusion	Long-range diffusion (through lattice)
Interface, diffusion or mixed control?	Interface control	Mixed control
Examples	Martensite Twinning Symmetric tilt boundary	Mainly inter-face control Mainly diffusion control

heat). Condensation and evaporation at a free solid surface are also included although they will not be treated in any depth²⁵.

While many transformations can be easily classified into the above system, there are other transformations where difficulties arise. For example, the bainite transformation takes place by thermally activated growth, but it also produces a shape change similar to that produced by the motion of a glissile interface. At present the exact nature of such transformations is unresolved.

There is a small class of transformations, known as *homogeneous transformations* that are not covered by Table 3.5. This is because they do not occur by the creation and migration of an interface, i.e. no nucleation stage is involved. Instead the transformation occurs *homogeneously* throughout the parent phase. Spinodal decomposition and certain ordering transformations are examples of this category and they will be discussed in Chapter 5.

3.5.1 Diffusion-Controlled and Interface-Controlled Growth²⁶

Let us now look more closely at the migration of an interface separating two phases of different composition. Consider for simplicity a β precipitate of almost pure B growing behind a planar interface into A-rich α with an initial composition X_0 as illustrated in Fig. 3.67. As the precipitate grows, the α adjacent to the interface becomes depleted of B so that the concentration of B in the α phase adjacent to the interface X_i decreases below the bulk concentration, Fig. 3.67a. Since growth of the precipitate requires a net flux of B atoms from the α to the β phase there must be a positive driving force across the interface $\Delta\mu_B$ as shown in Fig. 3.67b. The origin of this chemical potential difference can be seen in Fig. 3.67c. Clearly for growth to occur the interface composition must be greater than the equilibrium concentration X_e . By analogy with the migration of a high-angle grain boundary (Section 3.3.4) the net flux of B across the interface will produce an interface velocity v given by

$$v = M\Delta\mu_B^i/V_m \quad (3.49)$$

where M is the interface mobility and V_m is the molar volume of the β phase. The corresponding flux *across* the interface will be given by

$$J_B^i = -M\Delta\mu_B^i/V_m^2 \text{ moles of B m}^{-2} \text{ s}^{-1} \quad (3.50)$$

(The negative sign indicates that the flux is in the negative direction along the x -axis.) As a result of the concentration gradient in the α phase there will also be a flux of B atoms *towards* the interface J_B^a given by

$$J_B^a = -D \left(\frac{\partial C_B}{\partial x} \right)_{\text{interface}} \quad (3.51)$$

If a steady state exists at the interface these two fluxes must balance, i.e.

$$J_B^i = J_B^a \quad (3.52)$$

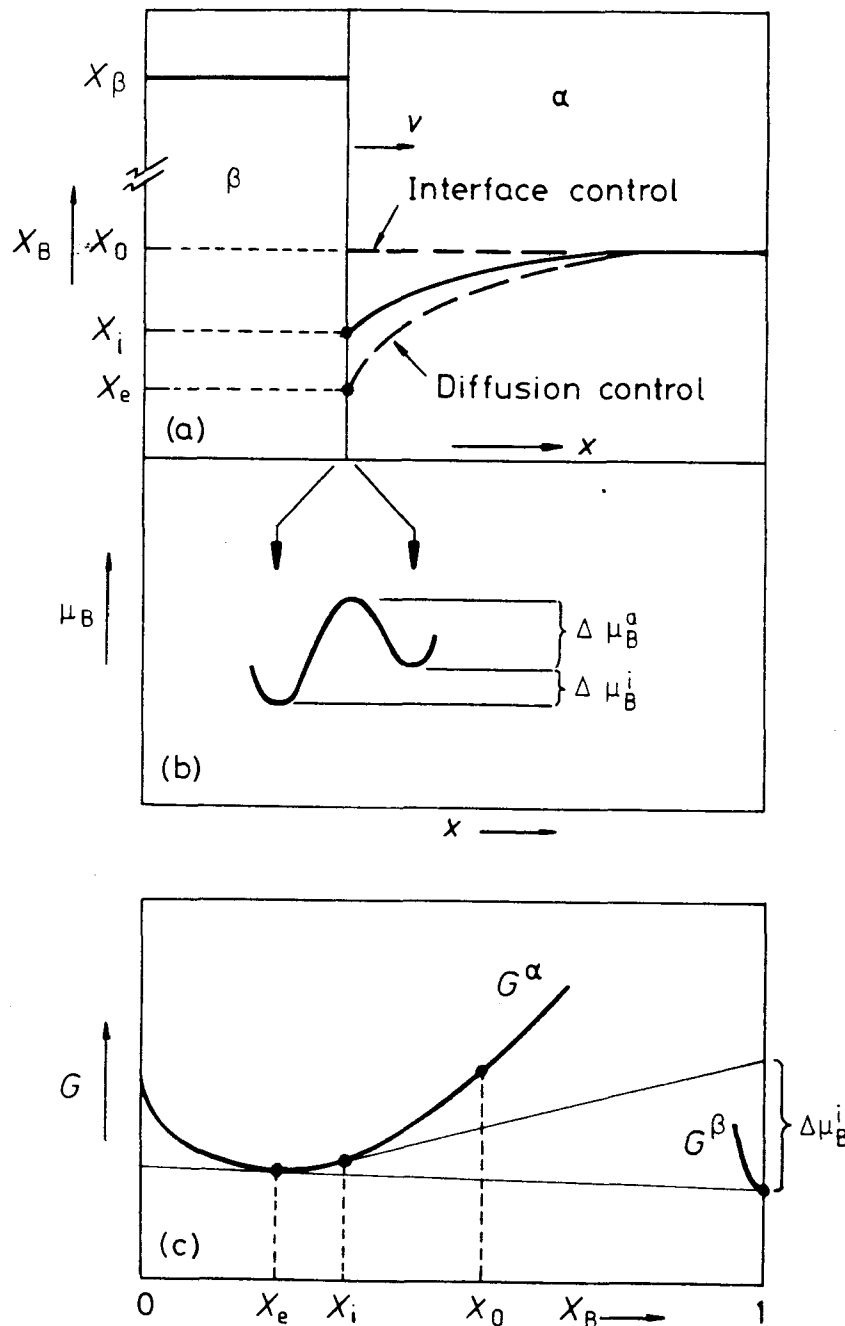


Fig. 3.67 Interface migration with long-range diffusion. (a) Composition profiles across the interface. (b) The origin of the driving force for boundary migration into the α phase. (c) A schematic molar free energy diagram showing the relationship between $\Delta \mu_B^i$, X_i and X_e . (Note that the solubility of A in the β phase is so low that the true shape of the free energy curve cannot be drawn on this scale.)

If the interface mobility is very high, e.g. an incoherent interface, $\Delta \mu_B^i$ can be very small and $X_i \approx X_e$. Under these circumstances there is effectively *local equilibrium* at the interface. The interface will then move as fast as diffusion allows, and growth takes place under *diffusion control*. The growth rate can then be evaluated as a function of time, say, by solving the diffusion equation with the boundary conditions $X_i = X_e$ and $X_B(\infty) = X_0$. Simple examples of this problem will be given in subsequent chapters in connection with solidification and diffusive transformations in solids.

When the interface has a lower mobility a greater chemical potential difference ($\Delta \mu_B^i$) is required to drive the interface reaction and there will be a departure from local equilibrium at the interface. The value of X_i that is chosen will be that which enables Equation 3.52 to be satisfied and the interface will then be migrating under *mixed control*. In the limit of a very low mobility it is possible that $X_i \approx X_0$ and $(\partial C / \partial x)_{\text{interface}}$ is almost zero. Under these conditions growth is said to be *interface controlled* and there is a maximum possible driving force $\Delta \mu_B^i$ across the interface.

It can easily be shown that for a dilute or ideal solution, the driving force $\Delta \mu_B^i$ is given by

$$\Delta \mu_B^i = RT \ln \frac{X_i}{X_e} \approx \frac{RT}{X_e} (X_i - X_e) \quad (3.53)$$

provided $(X_i - X_e) \ll X_e$ (see exercise 3.20). Thus the rate at which the interface moves under interface control should be proportional to the deviation of the interface concentration from equilibrium ($X_i - X_e$).

Let us now consider the question of why interface control should occur at all when the two phases have a different composition. At first sight it may appear that interface control should be very unlikely in practice. After all, the necessary long-range diffusion involves a great many atom jumps while the interface reaction essentially involves only one jump. Furthermore the activation energy for diffusion across the interface is not likely to be greater than for diffusion through the lattice—quite the contrary. On this basis, therefore, all interface reactions should be very rapid in comparison to lattice diffusion, i.e. all growth should be diffusion controlled. In many cases the above arguments are quite valid, but under certain conditions they are insufficient and may even be misleading.

Consider again the expression that was derived for the mobility of a high-angle grain boundary, Equation 3.22. A similar expression can be derived for the case of an interphase interface with $\Delta \mu_B^i$ replacing ΔG , (see exercise 3.19). It can be seen, therefore, that the above arguments neglect the effect of the accommodation factor (A), i.e. the probability that an atom crossing the boundary will be accommodated on arrival at the new phase. It is likely that incoherent interfaces and diffuse solid/liquid interfaces, as high-angle grain boundaries, will have values of A close to unity. These interfaces should therefore migrate under diffusion control. However, as will be demonstrated later, it is possible for certain types of coherent or semicoherent

interfaces, as well as smooth solid/liquid interfaces to have such low values of A that some degree of interface control is easily possible.

If two phases with different compositions, but the *same crystal structure* are separated by a coherent interface as shown in Fig. 3.32a, the interface can advance by the replacement of the α atoms in plane AA' with β atoms by normal lattice diffusion involving vacancies. There is no need for a separate interface reaction and the migration of this type of interface is therefore diffusion controlled. This situation arises during the growth of GP zones for example. The same arguments will apply if the interface is semicoherent provided the misfit dislocations can climb by vacancy creation or annihilation.

Quite a different situation arises when the two phases forming a coherent or semicoherent interface have *different crystal structures*. Consider for example the coherent close-packed interface between fcc and hcp crystals, Fig. 3.68a. If growth of the hcp phase is to occur by individual atomic jumps (i.e. so-called continuous growth) then an atom on a C site in the fcc phase must change into a B position as shown in Fig. 3.68b. It can be seen, however, that this results in a very high energy, unstable configuration with two atoms directly above each other on B sites. In addition a loop of Shockley partial dislocation is effectively created around the atom. An atom attempting such a jump will, therefore, be unstable and be forced back to its original position. The same situation will be encountered over the coherent regions of semicoherent interfaces separating phases with different crystal structures. Solid/vapour as well as smooth solid/liquid interfaces should behave in a similar manner, though perhaps to a lesser extent. If a single atom attaches itself to a flat close-packed interface it will raise the interfacial free energy and will therefore tend to detach itself again. It can thus be seen that continuous growth at the above type of interfaces will be very difficult, i.e. very low accommodation factors and low mobility are expected.

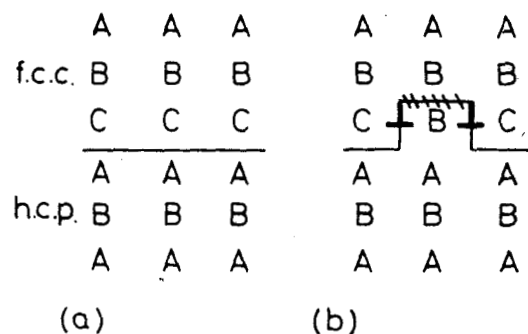


Fig. 3.68 Problems associated with the continuous growth of coherent interfaces between phases with different crystal structures. (After J.W. Martin and R.D. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, Cambridge, 1976.)

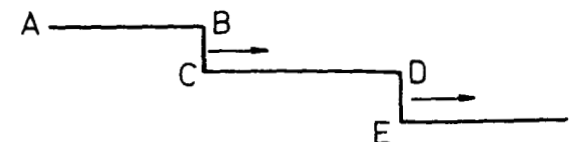


Fig. 3.69 The ledge mechanism.

A way of avoiding the difficulties of continuous growth encountered in the above cases is provided by the 'ledge' mechanism shown in Fig. 3.69. If the interface contains a series of ledges BC, DE normal to the facets AB, CD, EF, atoms will be able to transfer more easily across the ledges than the immobile facets and interface migration is therefore effected by the transverse migration of the ledges as shown.

Growth ledges have in fact been seen with the aid of the electron microscope on the surfaces of growing precipitates. For example Fig. 3.70 shows an electron micrograph and a schematic drawing of the growth ledges on an Mg_2Si plate in an Al-Mg-Si alloy²⁷. Note that growth ledges are usually hundreds of atom layers high.

When existing ledges have grown across the interface there is a problem of generating new ones. In Fig. 3.70 the source of new ledges is thought to be heterogeneous nucleation at the point of contact with another precipitate. The same problem will not be encountered if the precipitate is dissolving, however, since the edges of the plate will provide a continual source of ledges²⁸. It is thought that once nucleated, the rate at which ledges migrate across the planar facets should be diffusion controlled, i.e. controlled by how fast diffusion can occur to and from the ledges. However, the problem of nucleating new ledges may often lead to a degree of interface control on the overall rate at which the coherent or semicoherent interface can advance perpendicular to itself.

Growth ledges are by no means restricted to solid/solid systems. The first evidence for the existence of growth ledges came from studies of solid/vapour interfaces. They are also found on faceted solid/liquid interfaces.

The mechanism of interface migration can have important effects on the shape of second-phase inclusions. It was shown in Section 3.4.2 that in the absence of strain energy effects the *equilibrium shape* of a precipitate should be determined by the relative energies of the bounding interfaces. For example, a partially coherent precipitate should be disc or plate shaped with an aspect ratio of γ_i/γ_c where γ_i is the energy of the incoherent edges and γ_c is the energy of the coherent or semicoherent broad faces. However, the precipitate shape observed in practice may be prevented from achieving this equilibrium shape by the relative rates at which the coherent and incoherent interfaces can migrate. For example if there are problems of ledge nucleation the easier growth of the incoherent plate edges may lead to a larger aspect ratio than the equilibrium.

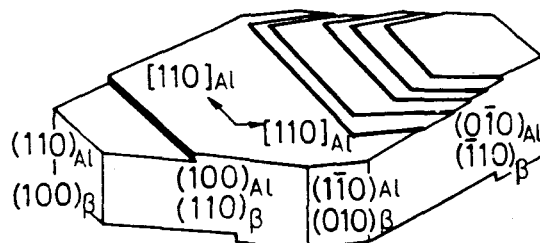


Fig. 3.70 (a) Growth ledges at an Mg_2Si plate in Al-1.5 wt% Mg_2Si , solution treated and aged 2 h at 350 °C. Dark field micrograph. (b) Schematic diagram of (a) showing ledges on Mg_2Si plate. (After G.C. Weatherly, *Acta Metallurgica*, 19 (1971) 181.)

References

1. This subject is covered in detail in J.W. Martin and R.D. Doherty *Stability of Microstructure in Metallic Systems*, Cambridge University Press, 1976, and in M. McLean 'Microstructural instabilities in metallurgical systems—a review' *Metal Science*, March 1978, p. 113.
2. Methods are described in J.M. Blakely, *Introduction to the properties of crystal surfaces*, Pergamon, Oxford, 1973, p. 53.

References

3. For a more detailed discussion, see J.W. Christian, *The Theory of Transformations in Metals and Alloys*, Second Edition, Part I, Pergamon, Oxford, 1975, p. 153.
4. R.S. Nelson, D.J. Mazey and R.S. Barnes, *Philosophical Magazine*, 11:91 (1965).
5. Sundquist, *Acta Metallurgica*, 12:67 (1964).
6. C.J. Simpson, K.T. Aust and W.C. Winegard, *Scripta Met.*, 3:171 (1969).
7. See for example J.W. Martin and R.W. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, 1976, p. 221.
8. J.W. Christian, *Metallurgical Transactions*, 21A: 799 (1990).
9. J.W. Christian, *Materials Science and Engineering*, A127: 215 (1990).
10. G.C. Weatherly and R.B. Nicholson, *Phil. Mag.* 17:801 (1968).
11. The interested reader is referred to M.I. Aaronson and J.K. Lee, 'Lectures on the theory of phase transformations', H.I. Aaronson (Ed.), Met. Soc. AIME, New York, 1975.
12. A quite general method for calculating the stress and strain fields in and around misfitting inclusions has been elucidated by J.D. Eshelby in *Proc. Roy. Soc., London*, A241 (1957) p. 376, and in *Progress in Solid Mechanics*, II, Chapter III (1961).
13. F.R.N. Nabarro, *Proc. Roy. Soc.* A175: 519 (1940).
14. Jong K. Lee and William C. Johnson, *Acta Metallurgica* 26:541-545 (1978).
15. R. Sankaran and C. Laird, *J. Mech. Phys. Solids*, 24:251-262 (1976) and *Phil. Mag.*, 29:179 (1974).
16. L.M. Brown and G.R. Woolhouse, *Phil. Mag.*, 21:329 (1970).
17. R. Sankaran and C. Laird, *Phil. Mag.*, 29:179 (1974).
18. K.E. Easterling and T. Johannesson, *Phil. Mag.*, 24:981 (1971).
19. A detailed treatment of solid/liquid interfaces can be found in D.P. Woodruff *The Solid-Liquid Interface*, Cambridge University Press, 1973.
20. K.A. Jackson, *Liquid Metals and Solidification*, ASM, Cleveland, Ohio (1958), p. 174.
21. W.A. Miller and G.A. Chadwick, *Acta Metallurgica*, 15:607 (1967).
22. W.A. Miller and G.A. Chadwick, *Proc. Roy. Soc.* A312:251 (1969).
23. J.W. Christian, *The Theory of Transformations in Metals and Alloys*, Edition 2, Part 1, Pergamon, 1975, p. 9.
24. J.W. Christian, 'Phase transformations in metals and alloys—an introduction', in *Phase Transformations*, 1, p. 1, Institute of Metallurgists, 1979.
25. The interested reader is referred to E. Rutner *et al.* (Eds.), *Condensation and Evaporation of Solids*, Gordon and Breach, New York, 1964.
26. For an advanced discussion of this subject see M. Hillert, 'Diffusion and interface control of reactions in alloys', *Met. Trans. A.*, 6A:5 (1975).
27. G.C. Weatherly, *Acta Metallurgica*, 19:181 (1971).

28. J.W. Martin and R.D. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, 1976, p. 17.

Further Reading

- J.M. Blakely, *Introduction to the Properties of Crystal Surfaces*, Pergamon Press, Oxford, 1973.
 G.A. Chadwick and D.A. Smith (Eds.), *Grain Boundary Structure and Properties*, Academic Press, London, 1976.
 H. Gleiter and B. Chalmers, 'High angle grain boundaries' in *Progress in Materials Science*, B. Chalmers, J.W. Christian and T.B. Massalski, (Eds.), Vol. 16, Pergamon, 1972.
 J.W. Martin and R.D. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, Cambridge, 1976.
 L.E. Murr, *Interfacial Phenomena in Metals and Alloys*, Addison-Wesley, London, 1975.
 D.P. Woodruff, *The Solid-Liquid Interface*, Cambridge University Press, Cambridge, 1973.

Exercises

- 3.1 Use the method of Section 3.1 to estimate the surface energy of {111}, {200} and {220} surface planes in an fcc crystal. Express your answer in J/surface atom and in J/m².
 3.2 Differentiate Equation 3.8 to obtain the slope of the $E_{SV} - \theta$ curve at $\theta = 0$.
 3.3 If a two-dimensional rectangular crystal is bounded by sides of lengths l_1 and l_2 show by differentiation that the equilibrium shape is given by

$$\frac{l_1}{l_2} = \frac{\gamma_2}{\gamma_1}$$

where γ_1 and γ_2 are the energies of the sides l_1 and l_2 respectively. (The area of the crystal $l_1 l_2$ is constant.)

- 3.4 (a) Measure θ for the low-angle tilt boundary in Fig. 3.11.
 (b) Determine the Burgers vector of the interface dislocations by making a Burgers circuit around one of the dislocations. Does the mean spacing of the dislocations agree with that predicted by Equation 3.9?
 3.5 Explain why grain boundaries move towards their centre of curvature during grain growth but away from their centre of curvature during recrystallization.
 3.6 (a) Suppose a recrystallized, dislocation-free grain is growing into a deformed matrix containing a dislocation density of 10^{16} m^{-2} (i.e. 10^{16} m/m^3). If the dislocations have an energy of $\mu b^2/4 \text{ J m}^{-1}$

calculate the pulling force acting on the recrystallized grain boundary. (Assume a shear modulus $\mu = 10^{10} \text{ N m}^{-2}$ and a Burgers vector $b = 0.28 \text{ nm}$.)

- (b) If the recrystallized grains grow from spherically shaped nuclei, what is the diameter of the smallest nucleus that can expand into the surrounding matrix? (Assume a grain boundary energy of 0.5 J m^{-2} .)
 3.7 Look up the equilibrium phase diagrams for the Al-Fe and Al-Mg systems. On the basis of these diagrams would you expect the grain boundary enrichment of Fe in dilute Al-Fe alloys to be greater or less than for Mg in dilute Al-Mg alloys at the same temperatures?
 3.8 Derive Equation 3.31.
 3.9 When a precipitate is surrounded by a spherical interface of radius r it is subjected to a pressure above that of the matrix by $2\gamma/r$. Consider a faceted precipitate with an equilibrium shape that of a square plate with a thickness of $2x_1$ and width $2x_2$. If the free energies of the broad faces and edges are respectively γ_1 and γ_2 , show that the broad faces exert a pressure on the precipitate (ΔP) given by

$$\Delta P = 2\gamma_2/x_2$$

(Hint: consider the total force acting on the periphery of the broad faces.) Show that the same result can be obtained by considering the pressure exerted by one of the edge faces of the plate.

- 3.10 Explain the structure and energies of coherent, semicoherent and incoherent interfaces, with particular reference to the role of orientation relationships and misfit.
 3.11 Fe-rich GP zones can form in dilute Al-Fe alloys. Given that the atomic radii are $1.43 \text{ \AA}$ for Al and $1.26 \text{ \AA}$ for Fe, would you expect the zones to be spherical or disc shaped?
 3.12 Mg can dissolve in Al to form a substitutional solid solution. Mg atoms are, however, bigger than Al atoms and each Mg atom therefore distorts the surrounding Al lattice, i.e. a coherency strain field effectively exists around each Mg atom. Using Equation 3.39 estimate the misfit strain energy. Express the answer in kJ mol^{-1} and eV atom^{-1} . (The shear modulus of Al = 25 GPa , the radius of an Al atom = $1.43 \text{ \AA}$, the radius of a Mg atom = $1.60 \text{ \AA}$.) What assumptions are implicit in this calculation?
 3.13 Explain why fully coherent precipitates tend to lose coherency as they grow.
 3.14 Show that the passage of a Shockley partial dislocation over every one of a given set of close-packed planes in fcc crystals produces a twin of the original crystal.
 3.15 If the ledges on the planar semicoherent interface in Fig. 3.69 move with a transverse velocity u what will be the overall velocity of the interface

- perpendicular to CD. Assume an infinite array of identical ledges of height (BC) = h and spacing (CD) = l .
- 3.16 Using arguments similar to those used in connection with Fig. 3.68 show that a coherent twin boundary in an fcc metal will not migrate by the random jumping of atoms across the interface. Suggest an interfacial structure that would result in a highly mobile interface (see exercise 3.15).
- 3.17 What are the most likely atomic processes involved in the migration of (i) solid/vapour interfaces, (ii) solid/liquid interfaces in non-metals, (iii) solid/liquid interfaces in metals.
- 3.18 By using a similar approach to the derivation of Equation 3.20 for a high-angle grain boundary, show that the net flux of B atoms across the α/β interface in Fig. 3.67 is given by

$$J_B^i = \frac{A_\beta n_\alpha v_\alpha}{RT} \exp\left(-\frac{\Delta\mu^a}{RT}\right) \Delta\mu_B^i$$

- 3.19 Derive Equation 3.53 for an ideal or dilute solution.
- 3.20 If an alloy containing β precipitates in an α matrix is given a solution treatment by heating to a temperature above the equilibrium β solvus the precipitates will dissolve. (See for example the phase diagram in Fig. 1.36.) Show with diagrams how the composition will change in the vicinity of an α/β interface during dissolution if the dissolution is (i) diffusion controlled, (ii) interface controlled, (iii) under mixed control. Indicate compositions by reference to a phase diagram where appropriate.

4

Solidification

Solidification and melting are transformations between crystallographic and non-crystallographic states of a metal or alloy. These transformations are of course basic to such technological applications as ingot casting, foundry casting, continuous casting, single-crystal growth for semiconductors, directionally solidified composite alloys, and more recently rapidly solidified alloys and glasses. Another important and complex solidification and melting process, often neglected in textbooks on solidification, concerns the process of fusion welding. An understanding of the mechanism of solidification and how it is affected by such parameters as temperature distribution, cooling rate and alloying, is important in the control of mechanical properties of cast metals and fusion welds. It is the objective of this chapter to develop some of the basic concepts of solidification, and apply these to some of the more important practical processes such as ingot casting, continuous casting and fusion welding. We then consider a few practical examples illustrating the casting or welding of engineering alloys in the light of the theoretical introduction.

4.1 Nucleation in Pure Metals

If a liquid is cooled below its equilibrium melting temperature (T_m) there is a driving force for solidification ($G_L - G_S$) and it might be expected that the liquid phase would spontaneously solidify. However, this is not always the case. For example under suitable conditions liquid nickel can be *undercooled* (or *supercooled*) to 250 K below T_m (1453 °C) and held there indefinitely without any transformation occurring. The reason for this behaviour is that the transformation begins by the formation of very small solid particles or *nuclei*. Normally undercoolings as large as 250 K are not observed, since in practice the walls of the liquid container and solid impurity particles in the liquid catalyse the nucleation of solid at undercoolings of only ~ 1 K. This is known as *heterogeneous* nucleation. The large undercoolings mentioned above are only obtained when no heterogeneous nucleation sites are available, i.e. when solid nuclei must form *homogeneously* from the liquid. Experimentally this can be achieved by dividing the liquid into tiny droplets, many of which remain impurity-free and do not solidify until very large undercoolings are reached¹.

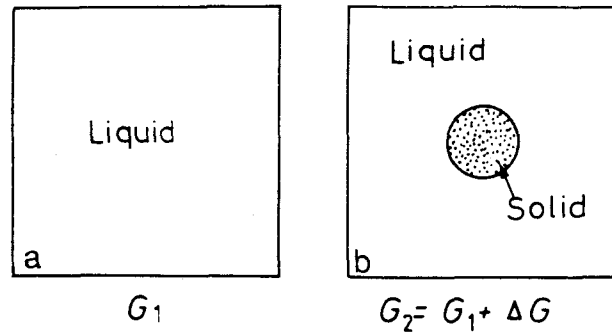


Fig. 4.1 Homogeneous nucleation.

4.1.1 Homogeneous Nucleation

Consider a given volume of liquid at a temperature ΔT below T_m with a free energy G_1 , Fig. 4.1a. If some of the atoms of the liquid cluster together to form a small sphere of solid, Fig. 4.1b, the free energy of the system will change to G_2 given by:

$$G_2 = V_S G_V^S + V_L G_V^L + A_{SL} \gamma_{SL}$$

where V_S is the volume of the solid sphere, V_L the volume of liquid, A_{SL} is the solid/liquid interfacial area, G_V^S and G_V^L are the free energies per unit volume of solid and liquid respectively, and γ_{SL} the solid/liquid interfacial free energy. The free energy of the system without any solid present is given by

$$G_1 = (V_S + V_L) G_V^L$$

The formation of solid therefore results in a free energy change $\Delta G = G_2 - G_1$ where:

$$\Delta G = -V_S \Delta G_V + A_{SL} \gamma_{SL} \quad (4.1)$$

and

$$\Delta G_V = G_V^L - G_V^S \quad (4.2)$$

For an undercooling ΔT , ΔG_V is given by Equation 1.17 as

$$\Delta G_V = \frac{L_v \Delta T}{T_m} \quad (4.3)$$

where L_v is the latent heat of fusion per unit volume. Below T_m , ΔG_V is positive so that the free energy change associated with the formation of a small volume of solid has a negative contribution due to the lower free energy of a bulk solid, but there is also a positive contribution due to the creation of a solid/liquid interface. The excess free energy associated with the solid particle can be minimized by the correct choice of particle shape. If γ_{SL} is

isotropic this is a sphere of radius r . Equation 4.1 then becomes

$$\Delta G_r = -\frac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \gamma_{SL} \quad (4.4)$$

This is illustrated in Fig. 4.2. Since the interfacial term increases as r^2 whereas the volume free energy released only increases as r^3 , the creation of small particles of solid always leads to a free energy increase. It is this increase that is able to maintain the liquid phase in a metastable state almost indefinitely at temperatures below T_m . It can be seen from Fig. 4.2 that for a given undercooling there is a certain radius, r^* , which is associated with a maximum excess free energy. If $r < r^*$ the system can lower its free energy by dissolution of the solid, whereas when $r > r^*$ the free energy of the system decreases if the solid grows. Unstable solid particles with $r < r^*$ are known as clusters or embryos while stable particles with $r > r^*$ are referred to as nuclei— r^* is known as the critical nucleus size. Since $dG = 0$ when $r = r^*$ the critical nucleus is effectively in (unstable) equilibrium with the surrounding liquid.

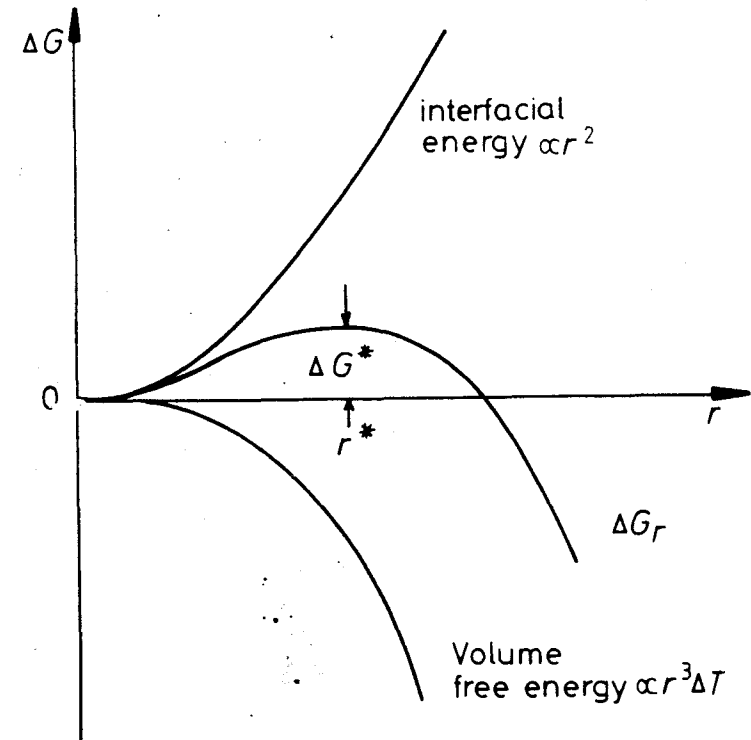


Fig. 4.2 The free energy change associated with homogeneous nucleation of a sphere of radius r .

It can easily be shown by differentiation of Equation 4.4 that

$$r^* = \frac{2\gamma_{SL}}{\Delta G_v} \quad (4.5)$$

and

$$\Delta G^* = \frac{16\pi\gamma_{SL}^3}{3(\Delta G_v)^2} \quad (4.6)$$

Substituting Equation 4.3 for ΔG_v gives

$$r^* = \left(\frac{2\gamma_{SL}T_m}{L_v} \right) \frac{1}{\Delta T} \quad (4.7)$$

and

$$\Delta G^* = \left(\frac{16\pi\gamma_{SL}^3 T_m^2}{3L_v^2} \right) \frac{1}{(\Delta T)^2} \quad (4.8)$$

Note how r^* and ΔG^* decrease with increasing undercooling (ΔT).

Equation 4.5 could also have been obtained from the Gibbs-Thomson equation. Since r^* is the radius of the solid sphere that is in (unstable) equilibrium with the surrounding liquid, the solidified sphere and liquid must then have the same free energy. From Equation 1.58 a solid sphere of radius r will have a free energy greater than that of bulk solid by $2\gamma V_m/r$ per mole or $2\gamma/r$ per unit volume. Therefore it can be seen from Fig. 4.3 that equality of

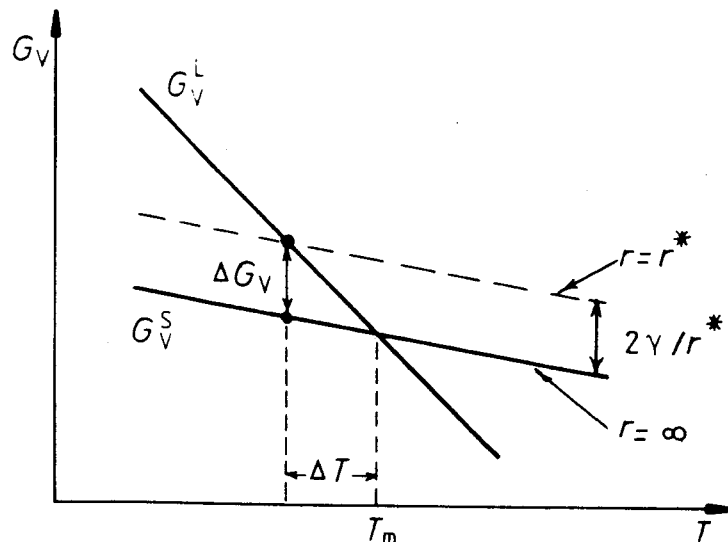


Fig. 4.3 Volume free energy as a function of temperature for solid and liquid phases, showing the origin of ΔG_v and r^* .

free energy implies

$$\Delta G_v = 2\gamma_{SL}/r^* \quad (4.9)$$

which is identical to Equation 4.5.

To understand how it is possible for a stable solid nucleus to form homogeneously from the liquid it is first necessary to examine the atomic structure of the liquid phase. From dilatometric measurements it is known that at the melting point the liquid phase has a volume 2–4% greater than the solid. Therefore there is a great deal more freedom of movement of atoms in the liquid and when averaged over a period of time the atom positions appear completely random. However, an instantaneous picture of the liquid would reveal the presence of many small close-packed clusters of atoms which are temporarily in the same crystalline array as in the solid, Fig. 4.4. On average the number of spherical clusters of radius r is given by

$$n_r = n_0 \exp \left(-\frac{\Delta G_r}{kT} \right) \quad (4.10)$$

where n_0 is the total number of atoms in the system, ΔG_r is the excess free energy associated with the cluster, Equation 4.4, and k is Boltzmann's constant. For a liquid above T_m this relationship applies for all values of r . Below T_m it only applies for $r \leq r^*$ because clusters greater than the critical size are stable nuclei of solid and no longer part of the liquid. Since n_r decreases exponentially with ΔG_r (which itself increases rapidly with r) the probability of finding a given cluster decreases very rapidly as the cluster size increases. For example by combining Equations 4.4 and 4.10 it can be shown (exercise 4.2) that 1 mm³ of copper at its melting point ($\sim 10^{20}$ atoms) should on average contain $\sim 10^{14}$ clusters of 0.3 nm radius (i.e. ~ 10 atoms) but only ~ 10 clusters with a radius of 0.6 nm (i.e. ~ 60 atoms). These numbers are of course only approximate. Such small clusters of atoms cannot be considered to be spherical, and even more important the effective value of γ used in calculating ΔG_r (equation 4.4) is very probably a function of the cluster size. However the above calculations do illustrate how sensitively cluster density depends on their size. Also, it can be seen that there is effectively a maximum

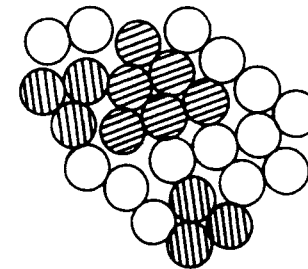


Fig. 4.4 A two-dimensional representation of an instantaneous picture of the liquid structure. Many close-packed crystal-like clusters (shaded) are present.

cluster size, ~ 100 atoms, which has a reasonable probability of occurring in the liquid. The same sort of calculations can be made at temperatures other than T_m . Below T_m there is an increasing contribution from ΔG_v in Equation 4.4 as the solid becomes progressively more stable and this has the effect of increasing the 'maximum' cluster size somewhat. Figure 4.5 shows schematically how $r_{\max}$ varies with ΔT . Of course larger clusters than $r_{\max}$ are possible in large enough systems or given sufficient time, but the probability of finding clusters only slightly larger than $r_{\max}$ is extremely small.

The critical nucleus size r^* is also shown in Fig. 4.5. It can be seen that at small undercoolings, r^* is so large that there will be virtually no chance of forming a stable nucleus. But as ΔT increases r^* and ΔG^* decrease, and for supercoolings of ΔT_N or greater there is a very good chance of some clusters reaching r^* and growing into stable solid particles. In the small droplet experiment, therefore, homogeneous nucleation should occur when the liquid is undercooled by $\sim \Delta T_N$.

The same conclusion can also be reached by an energy approach. The creation of a critical nucleus can be considered to be a thermally activated process, i.e. a solid-like cluster must be able to cross the nucleation barrier ΔG^* before it becomes a stable nucleus. Since the probability of achieving this energy is proportional to $\exp(-\Delta G^*/kT)$ nucleation will only become possible when ΔG^* is reduced below some critical value which can be shown to be $\sim 78 kT$ (see below).

4.1.2 The Homogeneous Nucleation Rate

Let us consider how fast solid nuclei will appear in the liquid at a given undercooling. If the liquid contains C_0 atoms per unit volume, the number of

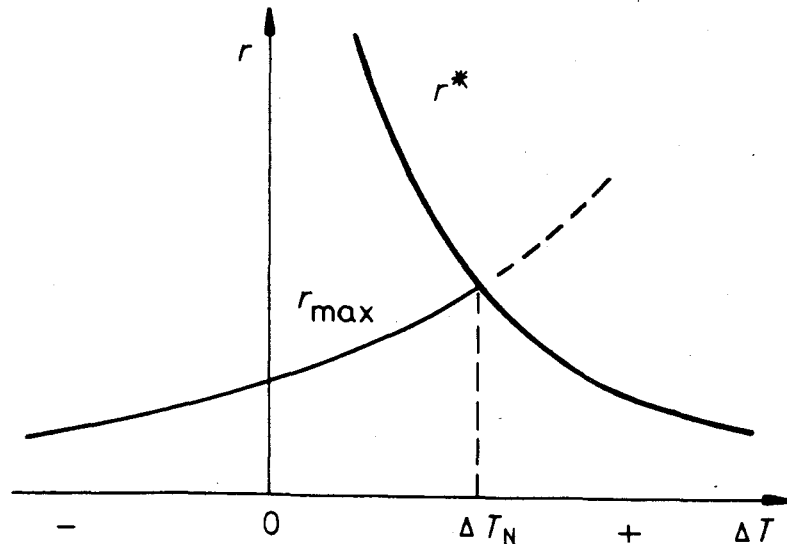


Fig. 4.5 The variation of r^* and $r_{\max}$ with undercooling ΔT .

clusters that have reached the critical size (C^*) can be obtained from Equation 4.10 as

$$C^* = C_0 \exp\left(-\frac{\Delta G_{\text{hom}}^*}{kT}\right) \text{ clusters m}^{-3} \quad (4.11)$$

The addition of one more atom to each of these clusters will convert them into stable nuclei and, if this happens with a frequency f_0 , the homogeneous nucleation rate will be given by

$$N_{\text{hom}} = f_0 C_0 \exp\left(-\frac{\Delta G_{\text{hom}}^*}{kT}\right) \text{ nuclei m}^{-3} \text{ s}^{-1} \quad (4.12)$$

where f_0 is a complex function that depends on the vibration frequency of the atoms, the activation energy for diffusion in the liquid, and the surface area of the critical nuclei. Its exact nature is not important here and it is sufficient to consider it a constant equal to $\sim 10^{11}$.^{*} Since C_0 is typically $\sim 10^{29}$ atoms m^{-3} a reasonable nucleation rate ($1 \text{ cm}^{-3} \text{ s}^{-1}$) is obtained when $\Delta G^* \sim 78 kT$.

$$N_{\text{hom}} = f_0 C_0 \exp\left\{-\frac{A}{(\Delta T)^2}\right\} \quad (4.13)$$

where A is relatively insensitive to temperature and is given by

$$A = \frac{16\pi\gamma_{\text{SL}}^3 T_m^2}{3L_v^2 kT}$$

N_{hom} is plotted as a function of ΔT in Fig. 4.6. As a result of the $(\Delta T)^2$ term, inside the exponential N_{hom} changes by orders of magnitude from essentially zero to very high values over a very narrow temperature range, i.e. there is effectively a critical undercooling for nucleation ΔT_N . This is the same as ΔT_N in Fig. 4.5, but Fig. 4.6 demonstrates more vividly how virtually no nuclei are formed until ΔT_N is reached after which there is an 'explosion' of nuclei.

The small droplet experiments of Turnbull *et al.*¹ have shown that ΔT_N is $\sim 0.2 T_m$ for most metals (i.e. $\sim 200 \text{ K}$). The measured values of ΔT_N have in fact been used along with Equation 4.13 to derive the values of interfacial free energy given in Table 3.4.

In practice homogeneous nucleation is rarely encountered in solidification.

^{*} Since atomic jumps from the liquid on to the cluster are thermally activated, f_0 will in fact diminish with decreasing temperature. In some metallic systems the liquid can be rapidly cooled to temperatures below the so-called glass transition temperature without the formation of crystalline solid. f_0 is very small at these temperatures and the supercooled liquid is a relatively stable metallic glass or amorphous metal. The variation of f_0 with temperature is very important with solid-state transformations, and it is covered in Chapter 5. For further details on alloys rapidly quenched from the melt see R.W. Cahn and P. Haasen (Eds), *Physical Metallurgy*, North-Holland, 1983, Chapter 28.

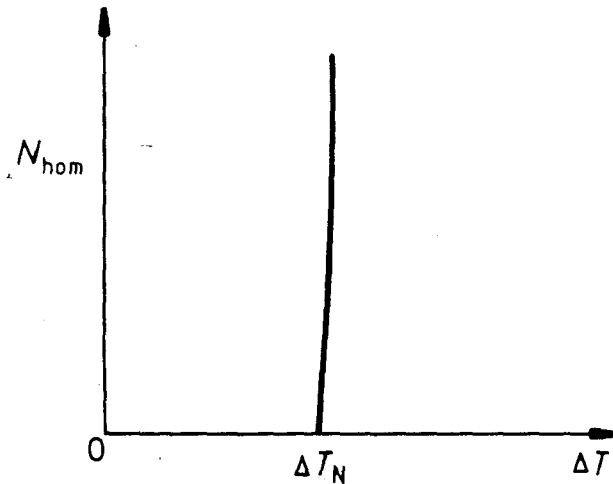


Fig. 4.6 The homogeneous nucleation rate as a function of undercooling ΔT . ΔT_N is the critical undercooling for homogeneous nucleation.

Instead heterogeneous nucleation occurs at crevices in mould walls, or at impurity particles suspended in the liquid.

4.1.3 Heterogeneous Nucleation

From the expression for ΔG^* (Equation 4.8) it can be seen that if nucleation is to be made easier at small undercoolings the interfacial energy term must be reduced. A simple way of effectively achieving this is if the nucleus forms in contact with the mould wall. Consider a solid embryo forming in contact with a perfectly flat mould wall as depicted in Fig. 4.7. Assuming γ_{SL} is isotropic it can be shown that for a given volume of solid the total interfacial energy of the system is minimized if the embryo has the shape of a spherical

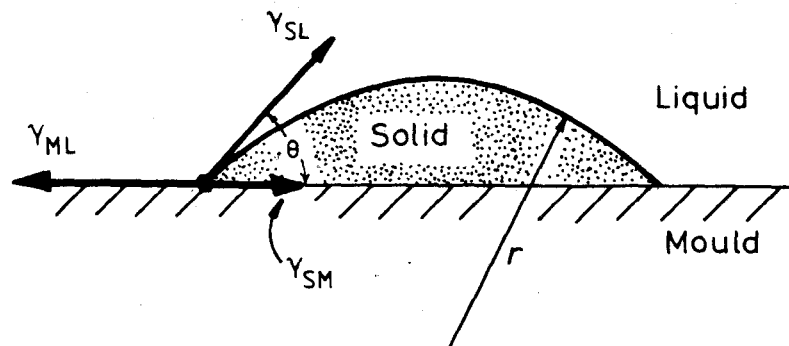


Fig. 4.7 Heterogeneous nucleation of spherical cap on a flat mould wall.

cap with a 'wetting' angle θ given by the condition that the interfacial tensions γ_{ML} , γ_{SM} and γ_{SL} balance in the plane of the mould wall.

$$\gamma_{ML} = \gamma_{SM} + \gamma_{SL} \cos \theta$$

or

$$\cos \theta = (\gamma_{ML} - \gamma_{SM}) / \gamma_{SL} \quad (4.14)$$

Note that the vertical component of γ_{SL} remains unbalanced. Given time this force would pull the mould surface upwards until the surface tension forces balance in all directions. Therefore Equation 4.14 only gives the optimum embryo shape on the condition that the mould walls remain planar.

The formation of such an embryo will be associated with an excess free energy given by

$$\Delta G_{het} = -V_S \Delta G_v + A_{SL} \gamma_{SL} + A_{SM} \gamma_{SM} - A_{SM} \gamma_{ML} \quad (4.15)$$

where V_S is the volume of the spherical cap, A_{SL} and A_{SM} are the areas of the solid/liquid and solid/mould interfaces, and γ_{SL} , γ_{SM} and γ_{ML} are the free energies of the solid/liquid, solid/mould and mould/liquid interfaces. Note that there are now three interfacial energy contributions. The first two are positive as they arise from interfaces created during the nucleation process. The third, however, is due to the destruction of the mould/liquid interface under the spherical cap and results in a negative energy contribution.

It can be easily shown (see exercise 4.6) that the above equation can be written in terms of the wetting angle (θ) and the cap radius (r) as

$$\Delta G_{het} = \left\{ -\frac{4}{3} \pi r^3 \Delta G_v + 4 \pi r^2 \gamma_{SL} \right\} S(\theta) \quad (4.16)$$

where

$$S(\theta) = (2 + \cos \theta)(1 - \cos \theta)^2 / 4 \quad (4.17)$$

Note that except for factor $S(\theta)$ this expression is the same as that obtained for homogeneous nucleation, Equation 4.4. $S(\theta)$ has a numerical value ≤ 1 dependent only on θ , i.e. the shape of the nucleus. It is therefore referred to as a shape factor. ΔG_{het} is shown in Fig. 4.8 along with ΔG_{hom} for comparison. By differentiation of Equation 4.16 it can be shown that

$$r^* = \frac{2\gamma_{SL}}{\Delta G_v} \quad (4.18)$$

and

$$\Delta G^* = \frac{16\pi\gamma_{SL}^3}{3\Delta G_v^2} \cdot S(\theta) \quad (4.19)$$

Therefore the activation energy barrier against heterogeneous nucleation (ΔG_{het}^*) is smaller than ΔG_{hom}^* by the shape factor $S(\theta)$. In addition the critical

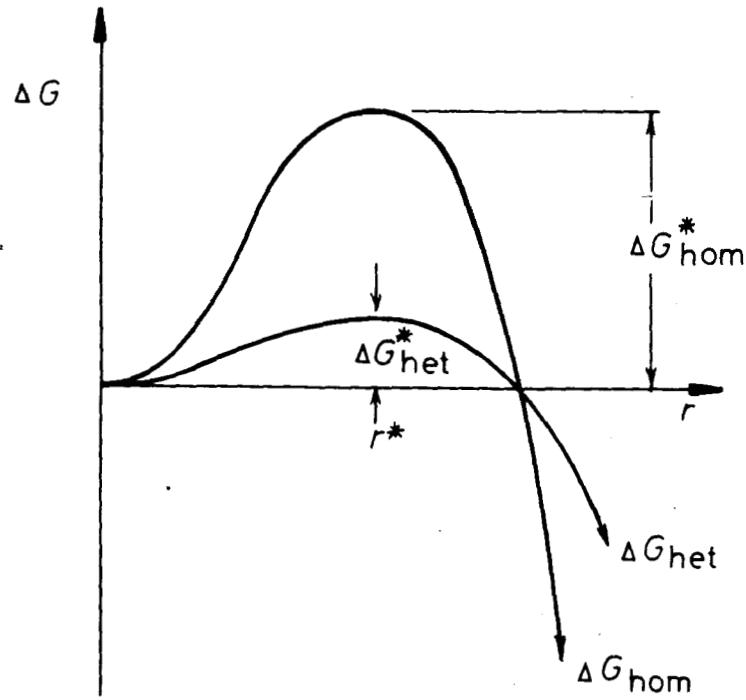


Fig. 4.8 The excess free energy of solid clusters for homogeneous and heterogeneous nucleation. Note r^* is independent of the nucleation site.

nucleus radius (r^*) is unaffected by the mould wall and only depends on the undercooling. This result was to be expected since equilibrium across the curved interface is unaffected by the presence of the mould wall.

Combining Equations 4.6 and 4.19 gives

$$\Delta G_{\text{het}}^* = S(\theta) \Delta G_{\text{hom}}^* \quad (4.20)$$

If for example $\theta = 10^\circ$, $S(\theta) \sim 10^{-4}$, i.e. the energy barrier for heterogeneous nucleation can be very much smaller than for homogeneous nucleation. Significant reductions are also obtained for higher values of θ , e.g. when $\theta = 30^\circ$, $S = 0.02$; even when $\theta = 90^\circ$, $S = 0.5$. It should be noted that the above model breaks down for $\theta = 0$. In this case the nucleus must be modelled in some other way, e.g. as shown in Fig. 4.12.

The effect of undercooling on ΔG_{het}^* and ΔG_{hom}^* is shown schematically in Fig. 4.9. If there are n_1 atoms in contact with the mould wall the number of nuclei should be given by

$$n^* = n_1 \exp\left(-\frac{\Delta G_{\text{het}}^*}{kT}\right) \quad (4.21)$$

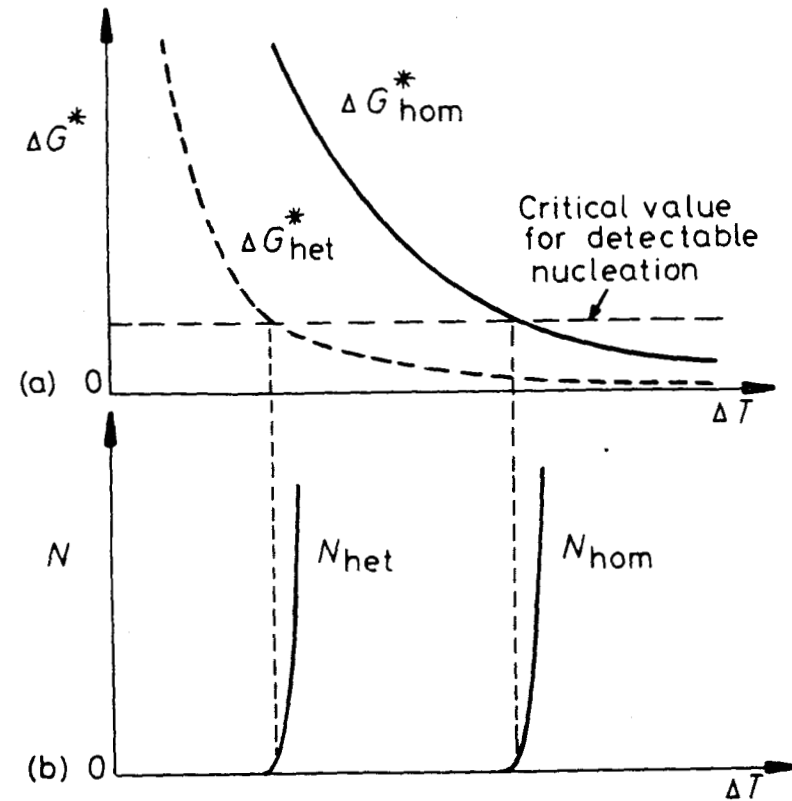


Fig. 4.9 (a) Variation of ΔG^* with undercooling (ΔT) for homogeneous and heterogeneous nucleation. (b) The corresponding nucleation rates assuming the same critical value of ΔG^* .

Therefore heterogeneous nucleation should become feasible when ΔG_{het}^* becomes sufficiently small. The critical value for ΔG_{het}^* should not be very different from the critical value for homogeneous nucleation. It will mainly depend on the magnitude of n_1 in the above equation. Assuming for the sake of simplicity that the critical value is again $\sim 78 kT$ it can be seen from Fig. 4.9 that heterogeneous nucleation will be possible at much lower undercoolings than are necessary for homogeneous nucleation.

To be more precise, the volume rate of heterogeneous nucleation ought to be given by an equation of the form

$$N_{\text{het}} = f_1 C_1 \exp\left(-\frac{\Delta G_{\text{het}}^*}{kT}\right) \quad (4.22)$$

where f_1 is a frequency factor similar to f_0 in Equation 4.12, C_1 is the number of atoms in contact with heterogeneous nucleation sites per unit volume of liquid.

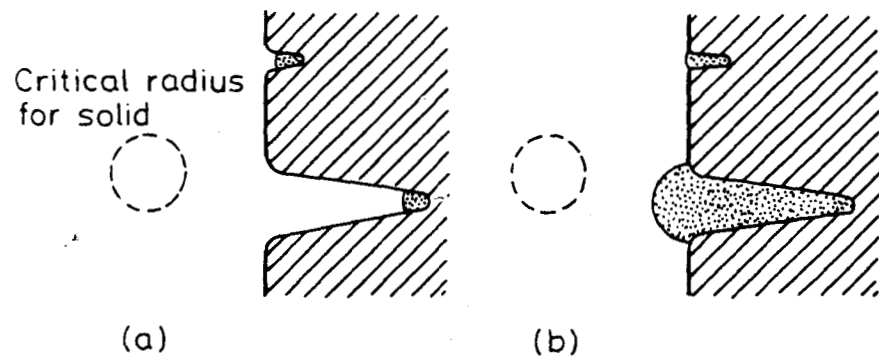


Fig. 4.10 Heterogeneous nucleation in mould-wall cracks. (a) The critical nuclei. (b) The upper nucleus cannot grow out of the crack while the lower one can. (After P.G. Shewmon, *Transformations in Metals*, © 1969 McGraw-Hill. Used with the permission of McGraw-Hill Book Company.)

So far it has been assumed that the mould wall is microscopically flat. In practice however it is likely to contain many microscopic cracks or crevices. It is possible to write down equations for the formation of a nucleus on such a surface but the result can be obtained more easily as follows. In both of the nucleation types considered so far it can be shown that

$$\Delta G^* = \frac{1}{2} V^* \Delta G_v \quad (4.23)$$

where V^* is the volume of the critical nucleus (sphere or cap). This equation, as well as Equation 4.7, are in fact quite generally true for any nucleation geometry. Thus, if a nucleus forms at the root of a crack the critical volume can be very small even if the wetting angle θ is quite large. Figure 4.10 shows an example where $\theta = 90^\circ$. Therefore nucleation from cracks or crevices should be able to occur at very small undercoolings even when the wetting angle θ is relatively large. Note however that for the crack to be effective the crack opening must be large enough to allow the solid to grow out without the radius of the solid/liquid interface decreasing below r^* .

In commercial practice heterogeneous nucleation is often enhanced by the addition of *inoculants* to the melt in order to refine the final grain size. The inoculating agent forms a solid compound with one of the components of the melt which then acts as a site for nucleation. According to the theory of heterogeneous nucleation outlined above the effectiveness of an inoculant should depend on the wetting angle and the surface roughness. Low values of θ are favoured by a low-energy interface between the inoculant and solid nucleus, γ_{SM} , which should in turn be favoured by good lattice matching between the particle and solid. However lattice matching alone is unable to account for the effectiveness of nucleants. Other contributing factors include chemical effects, as well as surface segregation and roughness. It is thus

difficult at present to predict the effectiveness of a given nucleant. In practice the aim of inoculant additions is of course not to reduce undercooling but to achieve a fine grain size, and then other variables such as the concentration of nucleating particles also becomes important.

4.1.4 Nucleation of Melting

Although nucleation during solidification usually requires some undercooling, melting invariably occurs at the equilibrium melting temperature even at relatively high rates of heating. This is due to the relative free energies of the solid/vapour, solid/liquid and liquid/vapour interfaces. It is always found that

$$\gamma_{SL} + \gamma_{LV} < \gamma_{SV} \quad (4.24)$$

Therefore the wetting angle $\theta = 0$ and no superheating is required for nucleation of the liquid. In fact this interfacial energy relationship implies that a thin liquid layer should even be able to form below T_m (see exercise 4.10). This phenomenon has not, however, been verified for metals as yet.

It is interesting to note that although T_m is a well-defined parameter in metallurgy, the actual atomic *mechanism* of melting is still not properly understood (for a good discussion of this phenomenon see, e.g. Cahn, 1978²). The solid $\rightarrow$ melt transformation in metals corresponds to an equivalent increase in vacancy concentration of as much as 10%, which is difficult to explain in the usual terms of defect structures. The melt, on this basis, might simply be considered to consist of an array of voids (condensed vacancies) surrounded by loose regions of disordered crystal (Frenkel's theory³). The sudden change from long-range crystallographic order to this loose, disordered structure may be associated with the creation of avalanches of dislocations which effectively break up the close-packed structure as melting occurs, as proposed by Cotterill *et al.* (1975)⁴ on the basis of computer simulation experiments. There are, however, problems of quantifying this dislocation mechanism with dilatometric observations, and a more refined theory of melting is awaited.

4.2 Growth of a Pure Solid

It was shown in Section 3.4.6 that there are basically two different types of solid/liquid interface: an atomically rough or diffuse interface associated with metallic systems, and an atomically flat or sharply defined interface often associated with non-metals. Because of the differences in atomic structure these two types of interface migrate in quite different ways. Rough interfaces

migrate by a *continuous growth* process while flat interfaces migrate by a *lateral growth* process involving ledges.

4.2.1 Continuous Growth

The migration of a diffuse solid/liquid interface can be treated in a similar way to the migration of a random high-angle grain boundary. The free energy of an atom crossing the S/L interface will vary as shown in Fig. 3.24 except one solid grain is replaced by the liquid phase. The activation energy barrier ΔG^a should be approximately the same as that for diffusion in the liquid phase, and the driving force for solidification (ΔG) will then be given by

$$\Delta G = \frac{L}{T_m} \cdot \Delta T_i \quad (4.25)$$

where L is the latent heat of melting and ΔT_i is the undercooling of the interface below the equilibrium melting temperature T_m . By analogy with Equation 3.21 therefore, the net rate of solidification should be given by an equation of the form

$$v = k_1 \Delta T_i \quad (4.26)$$

where k_1 has the properties of boundary mobility. A full theoretical treatment indicates that k_1 has such a high value that normal rates of solidification can be achieved with interfacial undercoolings (ΔT_i) of only a fraction of a degree Kelvin. For most purposes therefore ΔT_i can be ignored and the solid/liquid interface is assumed to be at the equilibrium melting temperature. In other words the solidification of metals is usually a diffusion controlled process. For pure metals growth occurs at a rate controlled by heat conduction (diffusion) whereas alloy solidification is controlled by solute diffusion.

The above treatment is applicable to diffuse interfaces where it can be assumed that atoms can be received at any site on the solid surface, i.e. the accommodation factor A in Equation 3.22 is approximately unity. For this reason it is known as continuous growth. Such a mode of growth is reasonable because the interface is disordered and atoms arriving at random positions on the solid will not significantly disrupt the equilibrium configuration of the interface. The situation is, however, more complex when the equilibrium interface structure is atomically smooth as in the case of many non-metals.

4.2.2 Lateral Growth

It will be recalled that materials with a high entropy of melting prefer to form atomically smooth, close-packed interfaces. For this type of interface the minimum free energy also corresponds to the minimum internal energy, i.e. a

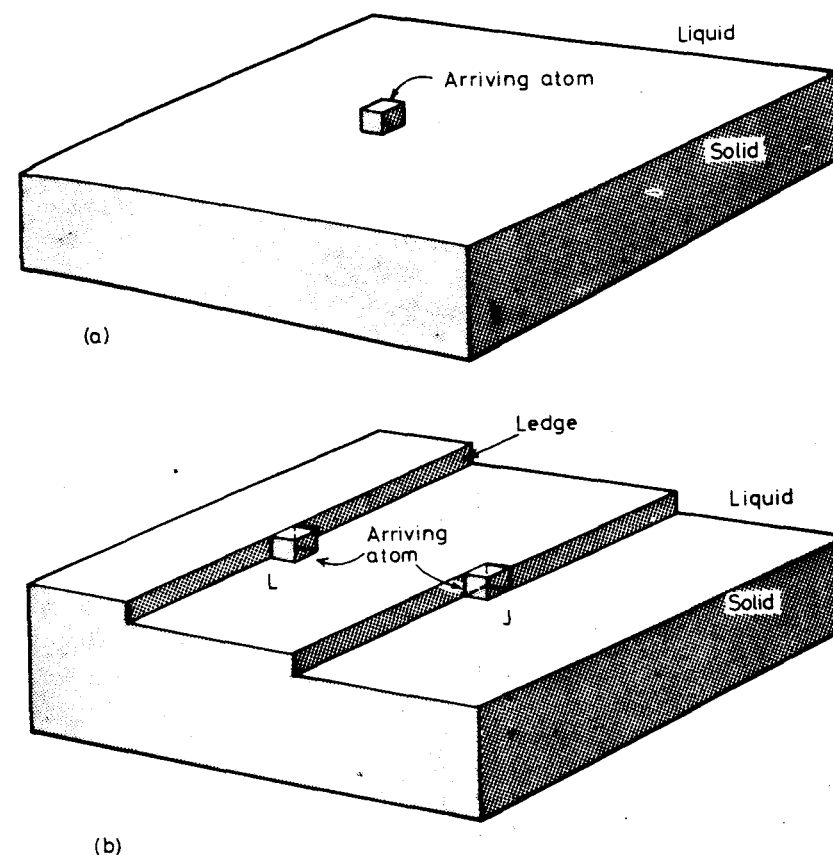


Fig. 4.11 Atomically smooth solid/liquid interfaces with atoms represented by cubes. (a) Addition of a single atom onto a flat interface increases the number of 'broken bonds' by four. (b) Addition to a ledge (L) only increases the number of broken bonds by two, whereas at a jog in a ledge (J) there is no increase.

minimum number of broken 'solid' bonds. If a single atom leaves the liquid and attaches itself to the flat solid surface, Fig. 4.11a, it can be seen that the number of broken bonds associated with the interface, i.e. the interfacial energy, will be increased. There is therefore little probability of the atom remaining attached to the solid and it is likely to jump back into the liquid. In other words, atomically smooth interfaces have inherently low accommodation factors. However, if the interface contains ledges, Fig. 4.11b, 'liquid' atoms will be able to join the ledges with a much lower resulting increase in interfacial energy. If the ledge contains a jog, J, atoms from the liquid can join the solid without any increase in the number of broken bonds and the interfacial energy remains unchanged. Consequently the probability of an atom remaining attached to the solid at these positions is much greater than for an atom joining a facet. Smooth solid/liquid interfaces can therefore be

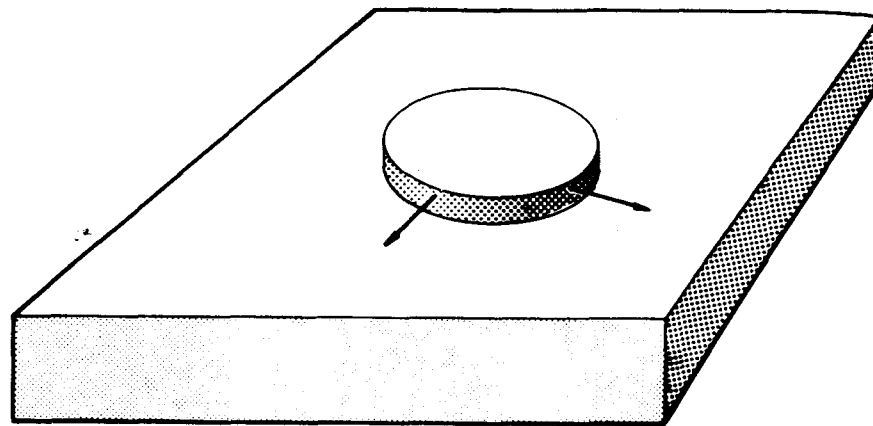


Fig. 4.12 Ledge creation by surface nucleation.

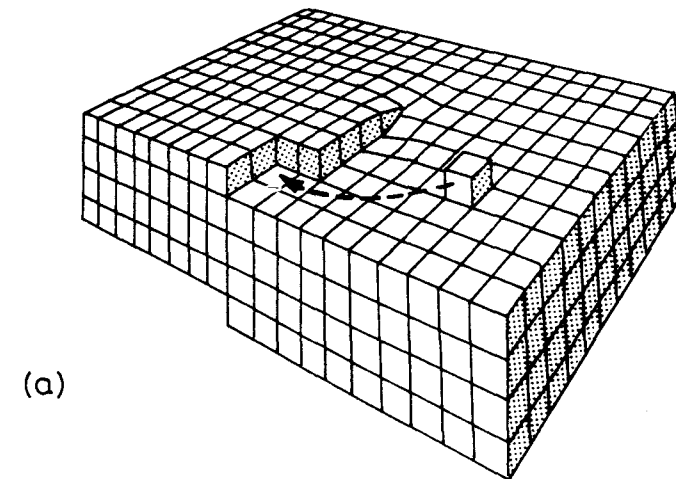
expected to advance by the lateral growth of ledges similar to that described for coherent solid/solid interfaces in Section 3.5.1. Since the ledges and jogs are a non-equilibrium feature of the interface, growth will be very dependent on how the ledges and jogs can be supplied. It is thought that there are basically three different ways in which this can be achieved. These are (i) by repeated surface nucleation, (ii) by spiral growth, and (iii) from twin boundaries.

Surface Nucleation

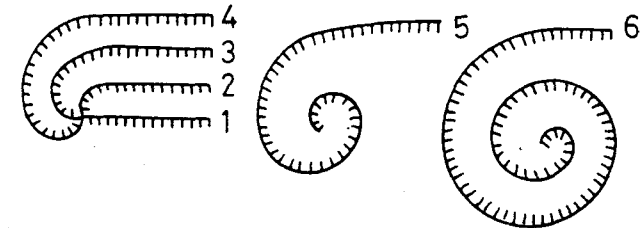
It was pointed out above that a single atom 'solidifying' on to a flat solid surface will be unstable and tend to rejoin the melt. However, if a sufficiently large number of atoms can come together to form a disc-shaped layer as shown in Fig. 4.12 it is possible for the arrangement to become self-stabilized and continue to grow. The problem of disc creation is the two-dimensional analogue of cluster formation during homogeneous nucleation. In this case the edges of the disc contribute a positive energy which must be counterbalanced by the volume free energy released in the process. There will therefore be a critical radius (r^*) associated with the two-dimensional nucleus which will decrease with increasing interface undercooling. Once nucleated the disc will spread rapidly over the surface and the rate of growth normal to the interface will be governed by the surface nucleation rate. A full theoretical treatment shows that

$$v \propto \exp(-k_2/\Delta T_i) \quad (4.27)$$

where k_2 is roughly constant. This is shown schematically in Fig. 4.14. Note that this mechanism is very ineffective at small undercoolings where r^* is very large.



(a)



(b)

Fig. 4.13 Spiral growth. (a) A screw dislocation terminating in the solid/liquid interface showing the associated ledge. (After W.T. Read Jr., *Dislocations in Crystals*, © 1953 McGraw-Hill. Used with the permission of McGraw-Hill Book Company.) Addition of atoms at the ledge causes it to rotate with an angular velocity decreasing away from the dislocation core so that a growth spiral develops as shown in (b). (After J.W. Christian, *The Theory of Phase Transformations in Metals and Alloys*, Pergamon Press, Oxford, 1965.)

Spiral Growth

If the solid contains dislocations that intersect the S/L interface the problem of creating new interfacial steps can be circumvented.

Consider for simplicity the introduction of a screw dislocation into a block of perfect crystal. The effect will be to create a step or ledge in the surface of the crystal as shown in Fig. 4.13a. The addition of atoms to the ledge will cause it to rotate about the point where the dislocation emerges, i.e. the ledge will never run out of the interface. If, on average, atoms add at an equal rate to all points along the step the angular velocity of the step will be initially greatest nearest to the dislocation core. Consequently as growth proceeds the ledge will develop into a growth spiral as shown in Fig. 4.13b. The spiral

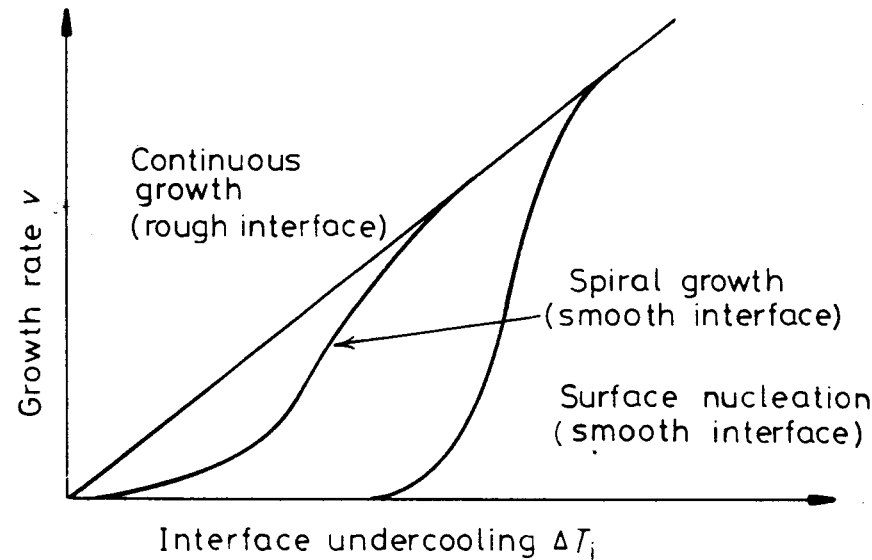


Fig. 4.14 The influence of interface undercooling (ΔT_i) on growth rate for atomically rough and smooth interfaces.

tightens until it reaches a minimum radius of curvature r^* at which it is in equilibrium with the surrounding liquid and can decrease no more. Further out the radius of curvature is less and the spiral can advance at a greater rate. Eventually a steady state is reached when the spiral appears to be rotating with a constant angular velocity. A complete theoretical treatment of this situation shows that for spiral growth the normal growth rate v and the undercooling of the interface ΔT_i are related by an expression of the type

$$v = k_3(\Delta T_i)^2 \quad (4.28)$$

where k_3 is a materials constant. This variation is shown in Fig. 4.14 along with the variations for continuous growth and two-dimensional nucleation. Note that for a given solid growth rate the necessary undercooling at the interface is least for the continuous growth of rough interfaces. For a given undercooling, faceted interfaces are much less mobile and it is to be expected that the spiral growth mechanism will normally be more important than repeated nucleation.

Growth from Twin Intersections

Another permanent source of steps can arise where two crystals in different orientations are in contact. In solidification it is quite common for materials showing faceting to solidify as two crystals in twin orientations. Interfacial facets will therefore intersect at the twin boundary which can act as a permanent source of new steps thereby providing an easy growth mechanism similar to the growth spiral mechanism.

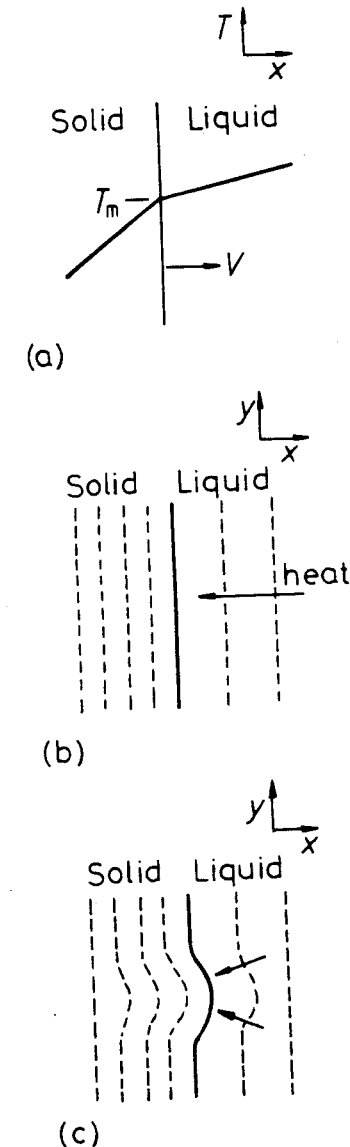


Fig. 4.15 (a) Temperature distribution for solidification when heat is extracted through the solid. Isotherms (b) for a planar S/L interface, and (c) for a protrusion.

4.2.3 Heat Flow and Interface Stability

In pure metals solidification is controlled by the rate at which the latent heat of solidification can be conducted away from the solid/liquid interface. Conduction can take place either through the solid or the liquid depending on the temperature gradients at the interface. Consider for example solid growing at

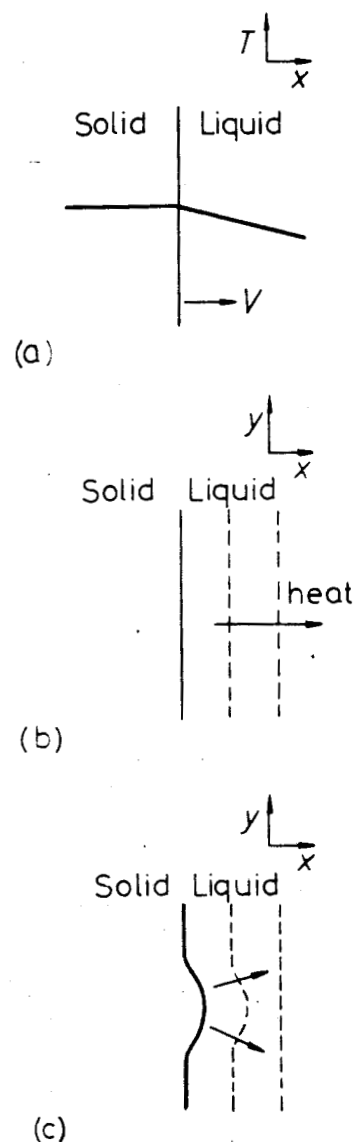


Fig. 4.16 As Fig. 4.15, but for heat conduction into the liquid.

a velocity v with a planar interface into a superheated liquid, Fig. 4.15a. The heat flow away from the interface through the solid must balance that from the liquid plus the latent heat generated at the interface, i.e.

$$K_S T'_S = K_L T'_L + v L_v \quad (4.29)$$

where K is the thermal conductivity, T' is the temperature gradient (dT/dx), the subscripts S and L stand for solid and liquid, v is the rate of growth of the

solid, and L_v is the latent heat of fusion per unit volume. This equation is quite general for a planar interface and even holds when heat is conducted into the liquid ($T'_L < 0$), Fig. 4.16a.

When a solid grows into a superheated liquid, a planar solid/liquid interface is stable. This can be shown as follows. Suppose that as a result of a local increase in v a small protrusion forms at the interface, Fig. 4.15a. If the radius of curvature of the protrusion is so large that the Gibbs-Thomson effect can be ignored the solid/liquid interface remains isothermal at essentially T_m . Therefore the temperature gradient in the liquid ahead of the nodule will

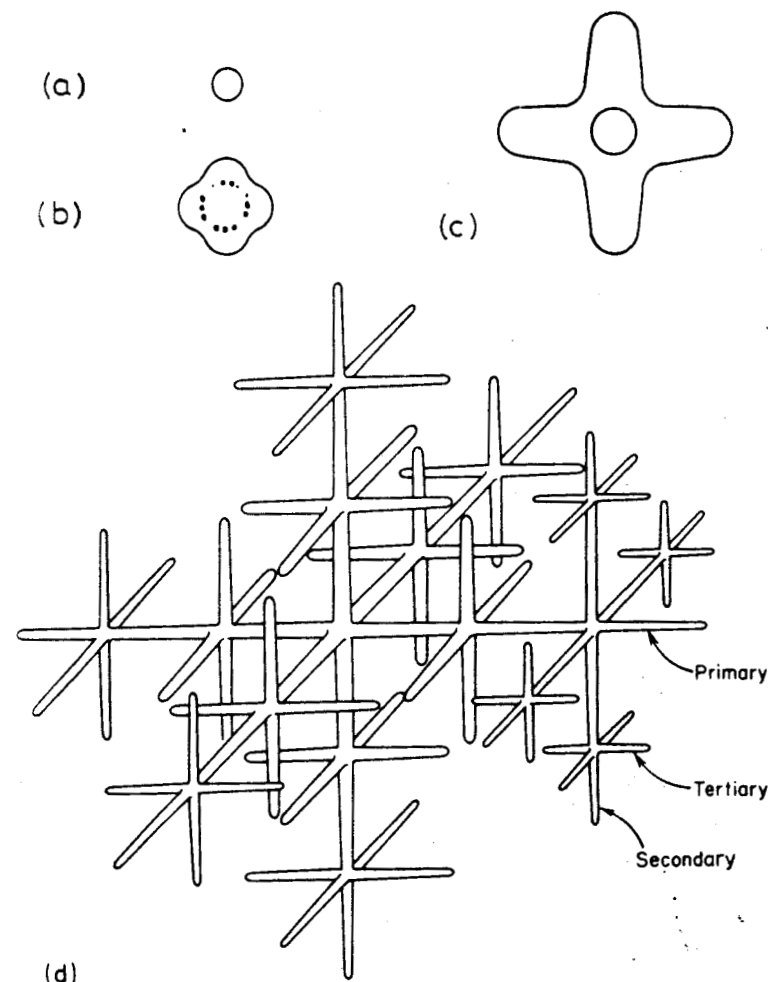


Fig. 4.17 The development of thermal dendrites: (a) a spherical nucleus; (b) the interface becomes unstable; (c) primary arms develop in crystallographic directions ($\{100\}$ in cubic crystals); (d) secondary and tertiary arms develop (after R.E. Reed-Hill, *Physical Metallurgy Principles*, 2nd. edn., Van Nostrand, New York, 1973).

increase while that in the solid decreases. Consequently more heat will be conducted into the protruding solid and less away so that the growth rate will decrease below that of the planar regions and the protrusion will disappear.

The situation is, however, different for a solid growing into supercooled liquid, Fig. 4.16. If a protrusion forms on the solid in this case the negative temperature gradient in the liquid becomes even more negative. Therefore heat is removed more effectively from the tip of the protrusion than from the surrounding regions allowing it to grow preferentially. A solid/liquid interface advancing into supercooled liquid is thus inherently unstable.

Heat conduction through the solid as depicted in Fig. 4.15, arises when solidification takes place from mould walls which are cooler than the melt. Heat flow *into* the liquid, however, can only arise if the liquid is supercooled below T_m . Such a situation can arise at the beginning of solidification if nucleation occurs at impurity particles in the bulk of the liquid. Since a certain supercooling is required before nucleation can occur, the first solid particles will grow into supercooled liquid and the latent heat of solidification will be conducted away into the liquid. An originally spherical solid particle will therefore develop arms in many directions as shown in Fig. 4.17. As the primary arms elongate their surfaces will also become unstable and break up into secondary and even tertiary arms. This shape of solid is known as a *dendrite*. *Dendrite* comes from the Greek for *tree*. Dendrites in pure metals are usually called *thermal dendrites* to distinguish them from dendrites in

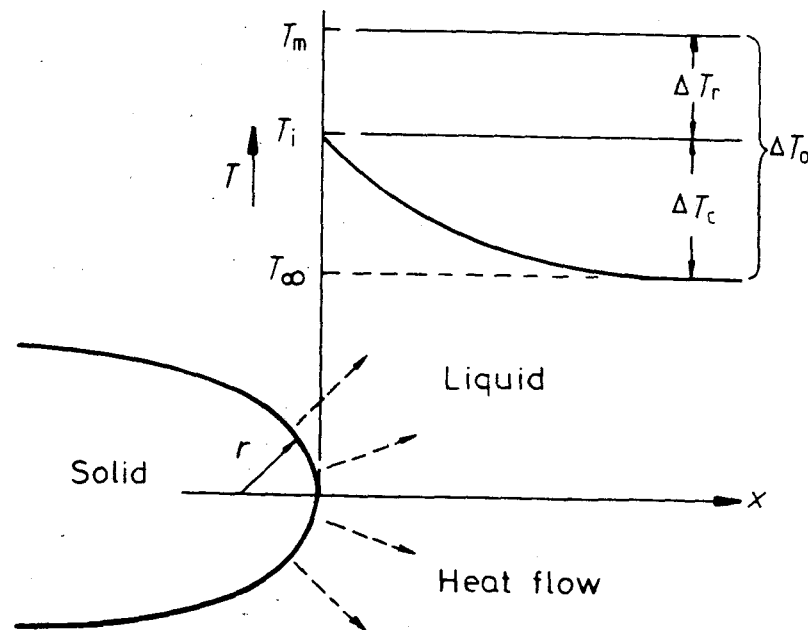


Fig. 4.18 Temperature distribution at the tip of a growing thermal dendrite.

alloys (see below). It is found experimentally that the dendrite arms are always in certain crystallographic directions: e.g. $\langle 100 \rangle$ in cubic metals, and $\langle 1100 \rangle$ in hcp metals⁵.

Let us now take a closer look at the tip of a growing dendrite. The situation is different from that of a planar interface because heat can be conducted away from the tip in three dimensions. If we assume the solid is isothermal ($T'_s = 0$) the growth rate of the tip v will be given by a similar equation to Equation 4.29 provided T'_L is measured in the direction of v . A solution to the heat-flow equations for a hemispherical tip shows that the (negative) temperature gradient T'_L is approximately given by $\Delta T_c/r$ where ΔT_c is the difference between the interface temperature (T_i) and the temperature of the supercooled liquid far from the dendrite (T_∞) as shown in Fig. 4.18. Equation 4.29 therefore gives

$$v = \frac{-K_L T'_L}{L_v} \approx \frac{K_L}{L_v} \cdot \frac{\Delta T}{r} \quad (4.30)$$

Thus for a given ΔT , rapid growth will be favoured by small values of r due to the increasing effectiveness of heat conduction as r diminishes. However ΔT is not independent of r . As a result of the Gibbs-Thomson effect equilibrium across a curved interface occurs at an undercooling ΔT_r below T_m given by

$$\Delta T_r = \frac{2\gamma T_m}{L_v r}$$

The minimum possible radius of curvature of the tip is when ΔT_r equals the total undercooling $\Delta T_0 = T_m - T_\infty$. This is just the critical nucleus radius r^* given by $(2\gamma T_m / L_v \Delta T_0)$. Therefore in general ΔT_r is given by $\Delta T_0 r^* / r$. Finally since $\Delta T_0 = \Delta T_c + \Delta T_r$ Equation 4.30 becomes

$$v = \frac{K_L}{L_v} \cdot \frac{1}{r} \left(1 - \frac{r^*}{r} \right) \quad (4.31)$$

It can thus be seen that the tip velocity tends to zero as $r \rightarrow r^*$ due to the Gibbs-Thomson effect and as $r \rightarrow \infty$ due to slower heat conduction. The maximum velocity is obtained when $r = 2r^*$.

4.3 Alloy Solidification

The solidification of pure metals is rarely encountered in practice. Even commercially pure metals contain sufficient impurities to change the characteristics of solidification from pure-metal to alloy behaviour. We now develop the theory a step further and examine the solidification of single-phase binary alloys. Following this we then consider the solidification of eutectic and peritectic alloys.

4.3.1 Solidification of Single-Phase Alloys

The alloys of interest in this section are those such as X_0 in Fig. 4.19. This phase diagram has been idealized by assuming that the solidus and liquidus are straight lines. It is useful to define a partition coefficient k as

$$k = \frac{X_S}{X_L} \quad (4.32)$$

where X_S and X_L are the mole fractions of solute in the solid and liquid in equilibrium at a given temperature. For the simple case shown in Fig. 4.19, k is independent of temperature.

The way in which such alloys solidify in practice depends in rather a complex way on temperature gradients, cooling rates and growth rates. Therefore let us simplify matters by considering the movement of a planar solid/liquid interface along a bar of alloy as shown in Fig. 4.20a. Such unidirectional solidification can be achieved in practice by passing the alloy in a crucible through a steep temperature gradient in a specially constructed furnace in which heat is confined to flow along the axis of the bar.

Let us examine three limiting cases:

1. Infinitely slow (equilibrium) solidification
2. Solidification with no diffusion in the solid but perfect mixing in the liquid
3. Solidification with no diffusion in the solid and only diffusional mixing in the liquid

Equilibrium Solidification

Alloy X_0 in Fig. 4.19 begins to solidify at T_1 with the formation of a small amount of solid with composition kX_0 . As the temperature is lowered more solid forms and, provided cooling is slow enough to allow extensive solid-state diffusion, the solid and liquid will always be homogeneous with compositions following the solidus and liquidus lines, Fig. 4.20b. The relative amounts of solid and liquid at any temperature are simply given by the lever rule. Note that, since solidification is one-dimensional, conservation of solute requires the two shaded areas in Fig. 4.20b to be equal (ignoring the differences in molar volume between the two phases). At T_3 the last drop of liquid will have a composition X_0/k and the bar of solid will have a composition X_0 along its entire length.

No Diffusion in Solid, Perfect Mixing in Liquid

Very often the rate of cooling will be too rapid to allow substantial diffusion in the solid phase. Therefore let us assume no diffusion takes place in the solid but that the liquid composition is kept homogeneous during solidification by efficient stirring. Again, assuming unidirectional solidification, the first solid will appear when the cooled end of the bar reaches T_1 in Fig. 4.21a, at which stage solid containing kX_0 mol of solute forms. Since $kX_0 < X_0$, this first

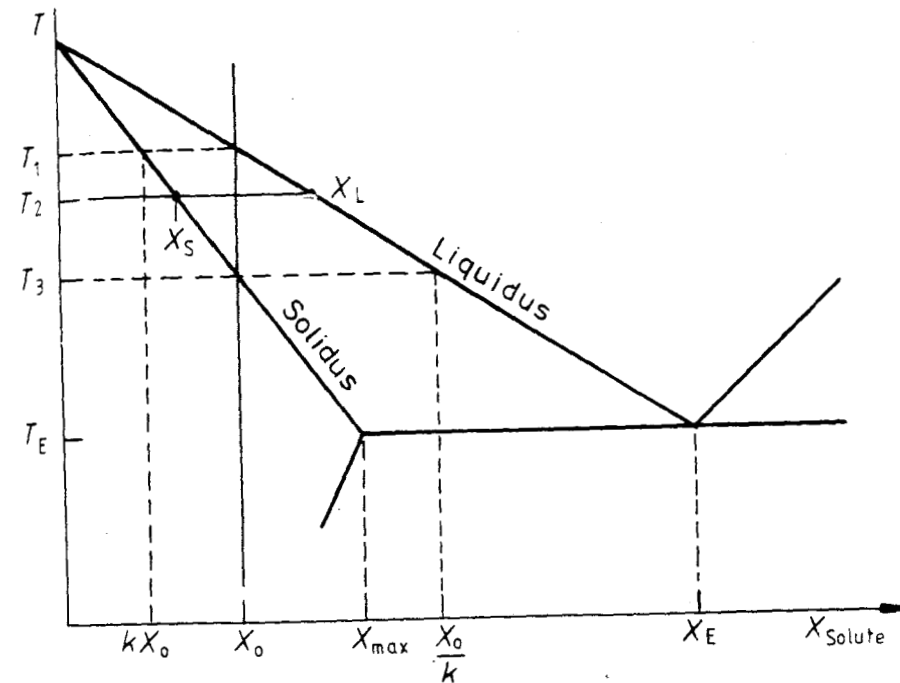


Fig. 4.19 A hypothetical phase diagram. $k = X_S/X_L$ is constant.

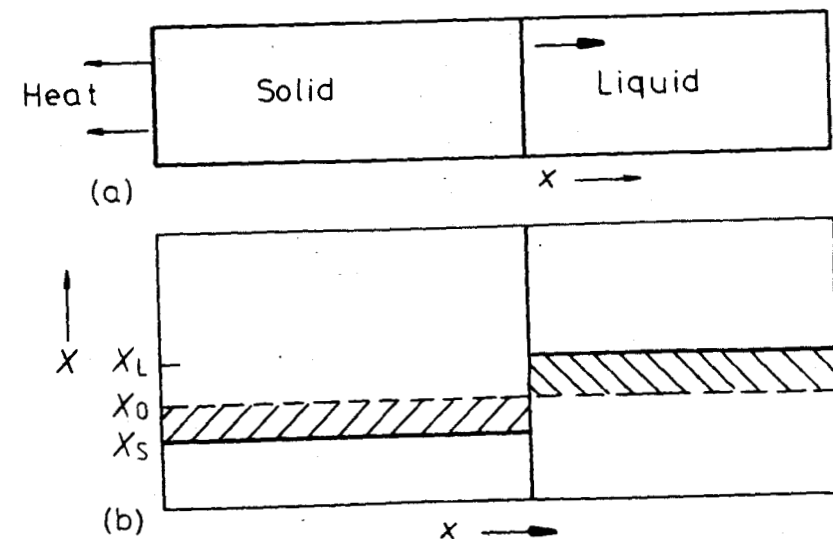


Fig. 4.20 Unidirectional solidification of alloy X_0 in Fig. 4.19. (a) A planar S/L interface and axial heat flow. (b) Corresponding composition profile at T_2 assuming complete equilibrium. Conservation of solute requires the two shaded areas to be equal.

solid will be purer than the liquid from which it forms so that solute is rejected into the liquid and raises its concentration above X_0 , Fig. 4.21b. The temperature of the interface must therefore decrease below T_1 before further solidification can occur, and the next layer of solid will be slightly richer in solute than the first. As this sequence of events continues the liquid becomes progressively richer in solute and solidification takes place at progressively lower temperatures, Fig. 4.21c. At any stage during solidification *local equilibrium* can be assumed to exist at the solid/liquid interface, i.e. for a given interface temperature the compositions of the solid and liquid in contact with one another will be given by the equilibrium phase diagram. However,

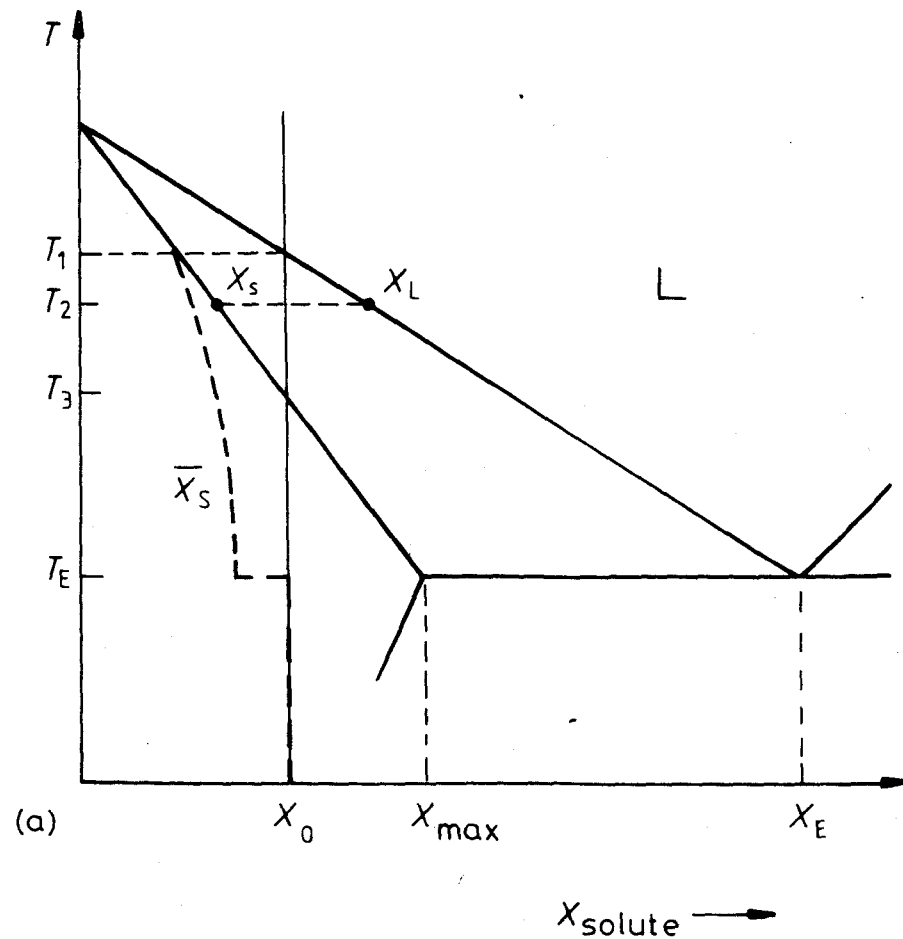


Fig. 4.21 Planar front solidification of alloy X_0 in Fig. 4.19 assuming no diffusion in the solid, but complete mixing in the liquid. (a) As Fig. 4.19, but including the mean composition of the solid. (b) Composition profile just under T_1 . (c) Composition profile at T_2 (compare with the profile and fraction solidified in Fig. 4.20b). (d) Composition profile at the eutectic temperature and below.

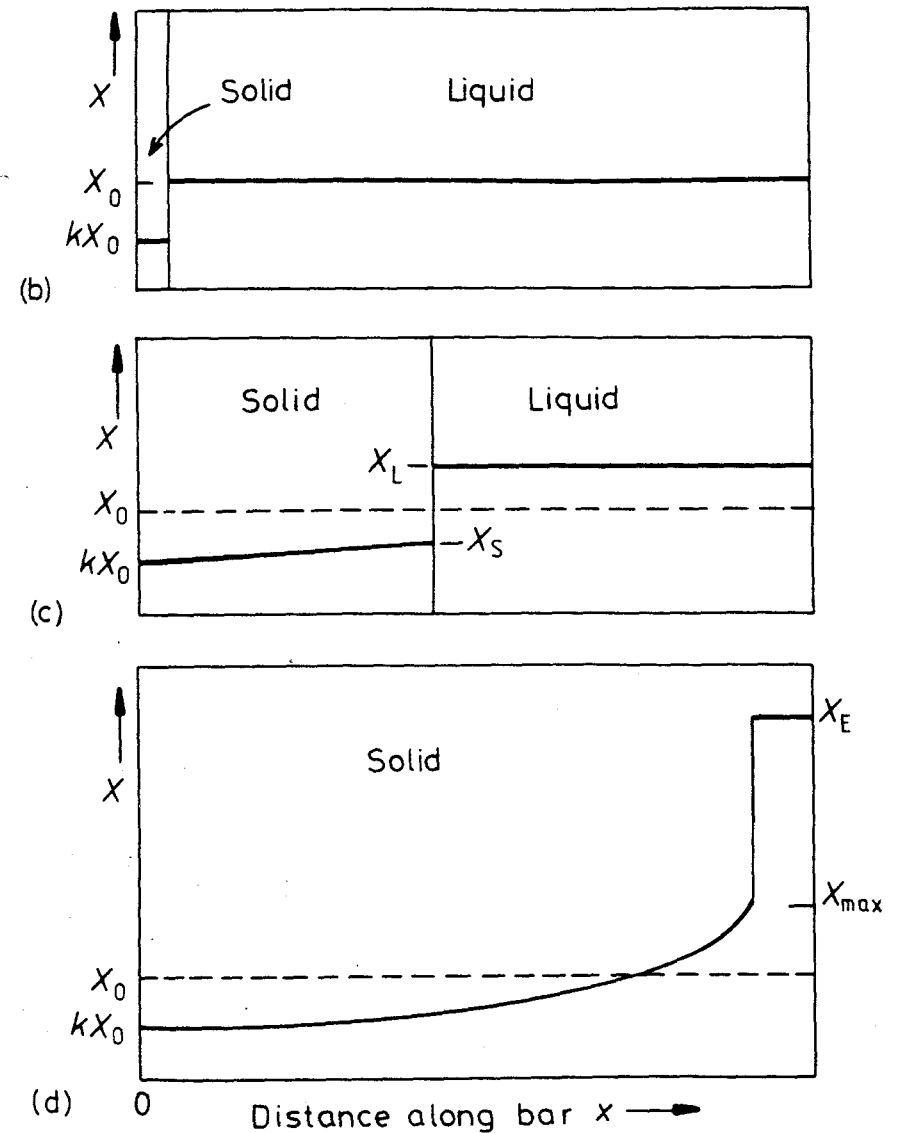


Fig. 4.21 (continued)

since there is no diffusion in the solid, the separate layers of solid retain their original compositions. Thus the mean composition of the solid ($\bar{X}_S$) is always lower than the composition at the solid/liquid interface, as shown by the dashed line in Fig. 4.21a. The relative amounts of solid and liquid for a given interface temperature are thus given by the lever rule using $\bar{X}_S$ and X_L . It follows that the liquid can become much richer in solute than X_0/k and it may even reach a eutectic composition, X_E , for example. Solidification will thus

tend to terminate close to T_E with the formation of a eutectic structure of $\alpha + \beta$. The completely solidified bar will then have a solute distribution as shown in Fig. 4.21d with $\bar{X}_S = X_0$.

The variation of X_S along the solidified bar can be obtained by equating the solute rejected into the liquid when a small amount of solid forms with the resulting solute increase in the liquid. Ignoring the difference in molar volume between the solid and liquid this gives

$$(X_L - X_S)df_S = (1 - f_S)dX_L$$

where f_S is the volume fraction solidified. Integrating this equation using the boundary condition $X_S = kX_0$ when $f_S = 0$ gives

$$X_S = kX_0(1 - f_S)^{(k-1)}$$

and

$$X_L = X_0 f_L^{(k-1)} \quad (4.33)$$

These equations are known as the *non-equilibrium lever rule* or the *Scheil equations*.

Note that for $k < 1$ these equations predict that when there is no diffusion in the solid there will always be some eutectic in the last drop of liquid to solidify, no matter how little solute is present. Also the equation is quite generally applicable even for non-planar solid/liquid interfaces provided the liquid composition is uniform and that the Gibbs-Thomson effect is negligible.

No Diffusion in Solid, Diffusional Mixing in Liquid

If there is no stirring or convection in the liquid phase the solute rejected from the solid will only be transported away by diffusion. Hence there will be a rapid build up of solute ahead of the solid and a correspondingly rapid increase in the composition of the solid formed, Fig. 4.22a. This is known as the initial transient. If solidification is made to occur at a constant rate, v , it can be shown that a steady state is finally obtained when the interface temperature reaches T_3 in Fig. 4.19⁶. The liquid adjacent to the solid then has a composition X_0/k and the solid forms with the bulk composition X_0 .

During steady-state growth the concentration profile in the liquid must be such that the rate at which solute diffuses down the concentration gradient away from the interface is balanced by the rate at which solute is rejected from the solidifying liquid, i.e.

$$-DC'_L = v(C_L - C_S) \quad (4.34)$$

where D is diffusivity in the liquid, C'_L stands for dC_L/dx at the interface, C_L and C_S are the solute concentrations of the liquid and solid in equilibrium at the interface (units: m^{-3}). Note the similarity of this equation to that describing the rate at which solidification occurs in pure metals, Equation 4.29.

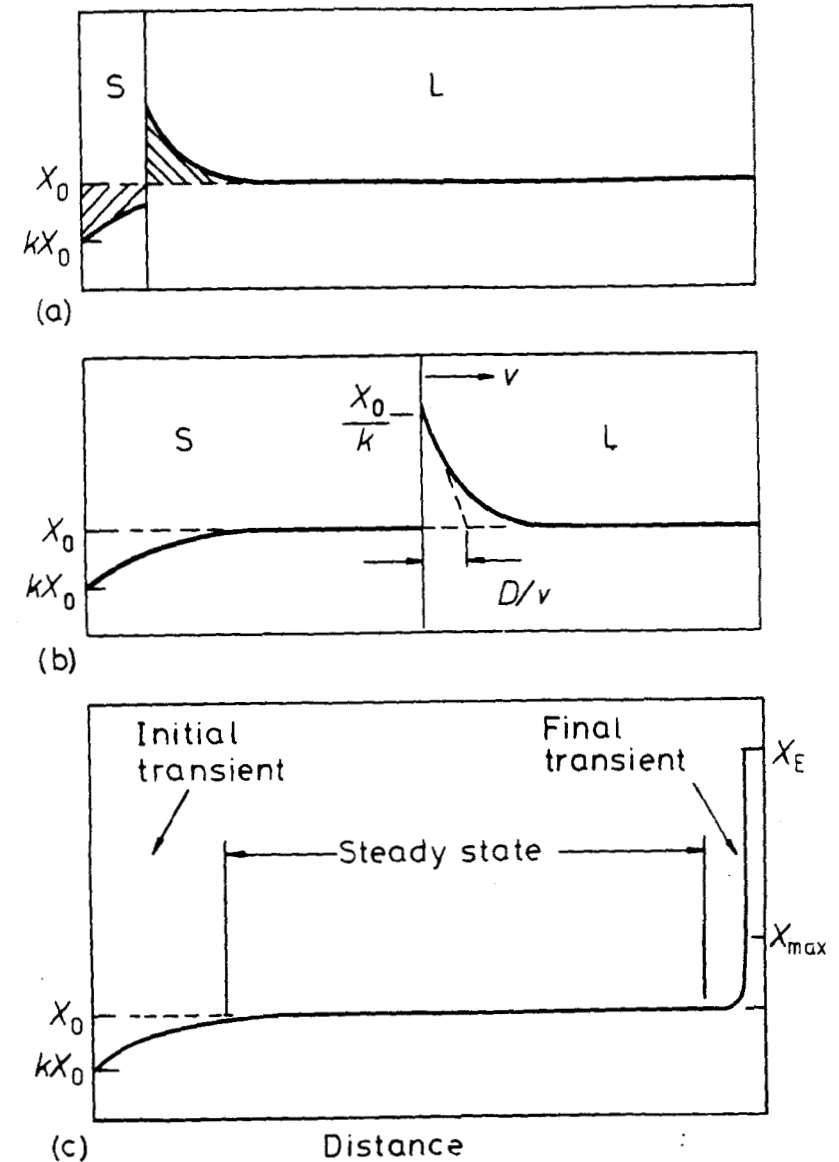


Fig. 4.22 Planar front solidification of alloy X_0 in Fig. 4.19 assuming no diffusion in the solid and no stirring in the liquid. (a) Composition profile when S/L interface temperature is between T_2 and T_3 in Fig. 4.19. (b) Steady-state solidification at T_3 . The composition solidifying equals the composition of the liquid far ahead of the solid (X_0). (c) Composition profile at T_E and below, showing the final transient.

If the diffusion equation is solved for steady-state solidification it can be shown that the concentration profile in the liquid is given by

$$X_L = X_0 \left\{ 1 + \frac{1-k}{k} \exp \left[-\frac{x}{(D/v)} \right] \right\} \quad (4.35)$$

i.e. X_L decreases exponentially from X_0/k at $x = 0$, the interface, to X_0 at large distances from the interface. The concentration profile has a characteristic width of D/v .

When the solid/liquid interface is within $\sim D/v$ of the end of the bar the bow-wave of solute is compressed into a very small volume and the interface composition rises rapidly leading to a final transient and eutectic formation, Fig. 4.22c.

In practice alloy solidification will usually possess features from all three of the cases discussed above. There will usually be some stirring either due to liquid turbulence caused by pouring, or because of convection currents, or gravity effects. However, stirring will not usually be sufficiently effective to prevent the formation of a boundary layer and some liquid diffusion will therefore be involved. Partial stirring does, however, have the effect of reducing the boundary layer thickness. The concentration profiles found in practice may thus exhibit features between those shown in Fig. 4.21d and 4.22c. In many cases diffusion in the solid must also be taken into account, e.g. when interstitial atoms or bcc metals are involved. In this case solute can diffuse away from the solidifying interface back into the solid as well as into the liquid, with the result that after solidification the alloy is more homogeneous.

Even when solidification is not unidirectional the above ideas can still often be applied at a microscopic level as will be discussed below. Unidirectional solidification has commercial importance in, for example, the production of creep resistant aligned microstructures for gas turbine blades. It is also used in the production of extremely pure metals (zone refining)⁷.

Cellular and Dendritic Solidification

So far we have considered solidification in which the growth front is planar. However, the diffusion of solute into the liquid during solidification of an alloy is analogous to the conduction of latent heat into the liquid during the solidification of a pure metal. At first sight therefore it would seem that the planar front should break up into dendrites. The problem is complicated, however, by the possibility of temperature gradients in the liquid.

Consider steady-state solidification at a planar interface as shown in Fig. 4.23. As a result of the varying solute concentration ahead of the solidification front there is a corresponding variation of the equilibrium solidification temperature, i.e. the liquidus temperature, as given by the line T_e in Fig. 4.23b. However, apart from the temperature of the interface, which is fixed by local equilibrium requirements, the actual temperature of the liquid

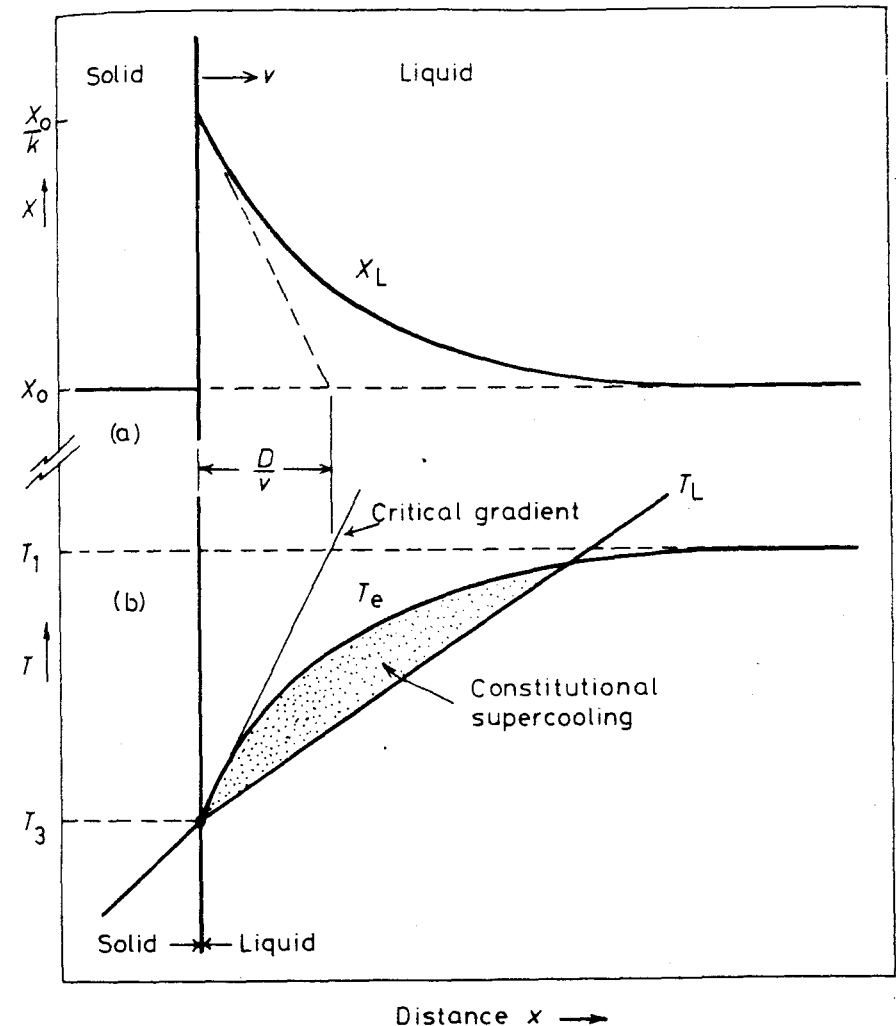


Fig. 4.23 The origin of constitutional supercooling ahead of a planar solidification front. (a) Composition profile across the solid/liquid interface during steady-state solidification. The dashed line shows dX_L/dx at the S/L interface. (b) The temperature of the liquid ahead of the solidification front follows line T_L . The equilibrium liquidus temperature for the liquid adjacent to the interface varies as T_e . Constitutional supercooling arises when T_L lies under the critical gradient.

can follow any line such as T_L . At the interface $T_L = T_e = T_3$ (defined in Fig. 4.19). If the temperature gradient is less than the critical value shown in Fig. 4.23b the liquid in front of the solidification front exists below its equilibrium freezing temperature, i.e. it is supercooled. Since the supercooling arises from compositional, or constitutional effects it is known as *constitutional supercooling*.

A necessary condition for the formation of stable protrusions on a planar interface is that there must exist a region of constitutional supercooling in the liquid. Assuming the T_L variation in Fig. 4.23b the temperature at the tip of any protrusion that forms will be higher than that of the surrounding interface. (In contrast to pure metals the interface in alloys need not be isothermal.) However, provided the tip remains below the local liquidus temperature (T_L) solidification is still possible and the protrusion can develop. On the other hand if the temperature gradient ahead of the interface is steeper than the critical gradient in Fig. 4.23b the tip will be raised above the liquidus temperature and the protrusion will melt back.

Under steady-state growth the critical gradient can be seen from Fig. 4.23 to be given by $(T_1 - T_3)/(D/v)$ where T_1 and T_3 are the liquidus and solidus temperatures for the bulk composition X_0 , Fig. 4.19. The condition for a stable planar interface is therefore

$$T'_L > \frac{(T_1 - T_3)}{(D/v)}$$

where T'_L stands for (dT_L/dx) at the interface. Or, regrouping the experimentally adjustable parameters T'_L and v , the condition for no constitutional supercooling is

$$(T'_L/v) > (T_1 - T_3)/D \quad (4.36)$$

$(T_1 - T_3)$ is known as the *equilibrium freezing range* of the alloy. Clearly planar front solidification is most difficult for alloys with a large solidification range and high rates of solidification. Except under well-controlled experimental conditions alloys rarely solidify with planar solid/liquid interfaces. Normally the temperature gradients and growth rates are not individually controllable but are determined by the rate at which heat is conducted away from the solidifying alloy.

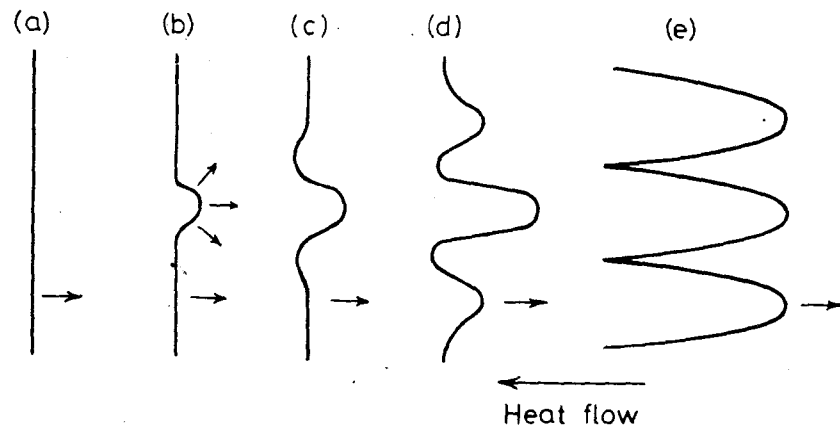


Fig. 4.24 The breakdown of an initially planar solidification front into cells.

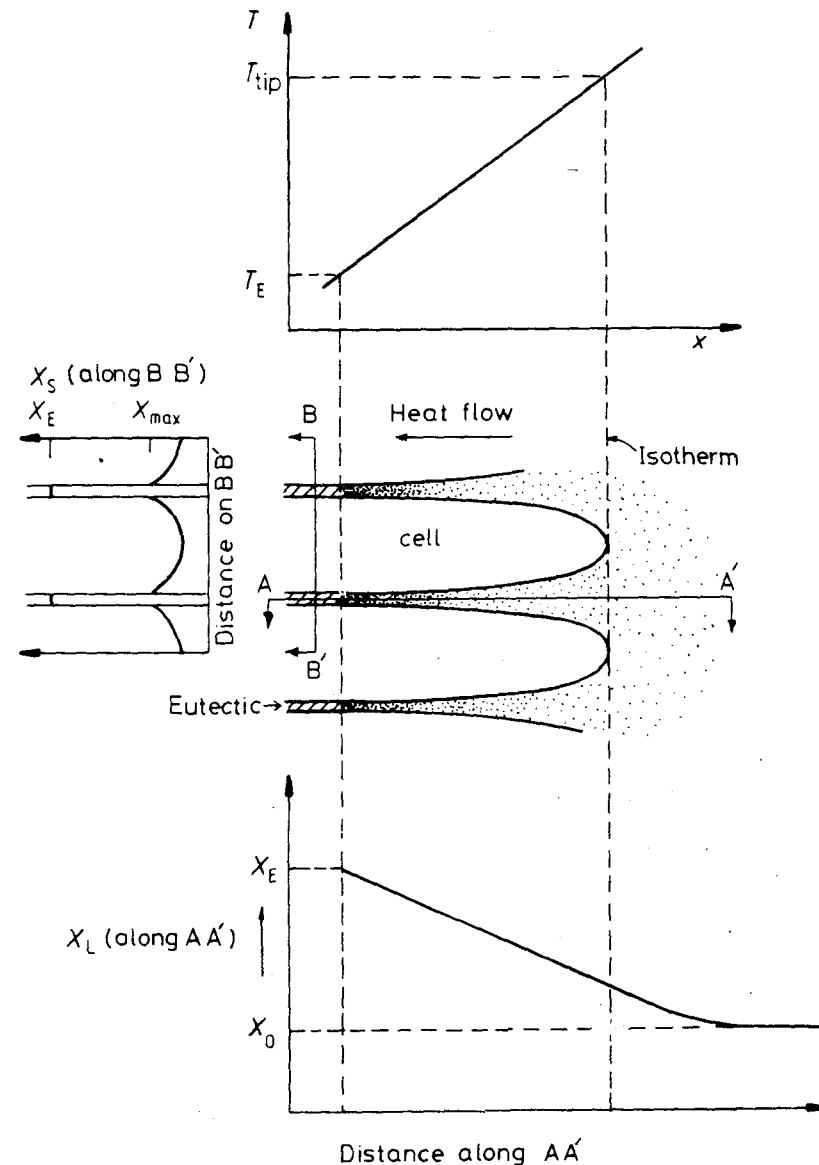


Fig. 4.25 Temperature and solute distributions associated with cellular solidification. Note that solute enrichment in the liquid between the cells, and coring in the cells with eutectic in the cell walls.

If the temperature gradient ahead of an initially planar interface is gradually reduced below the critical value the first stage in the breakdown of the interface is the formation of a cellular structure, Fig. 4.24. The formation of the first protrusion causes solute to be rejected laterally and pile up at the root of the protrusion (b). This lowers the equilibrium solidification temperature

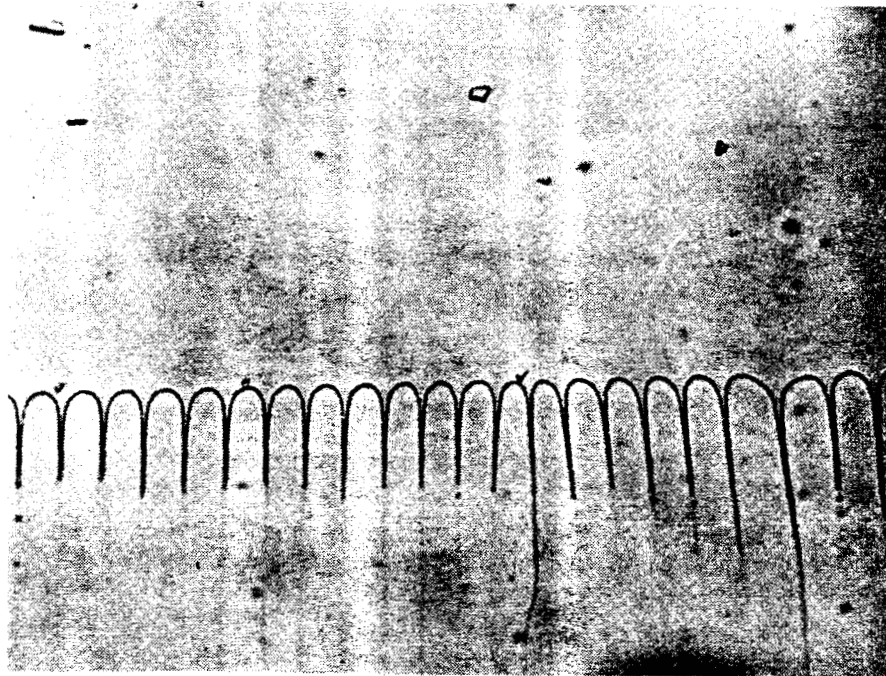
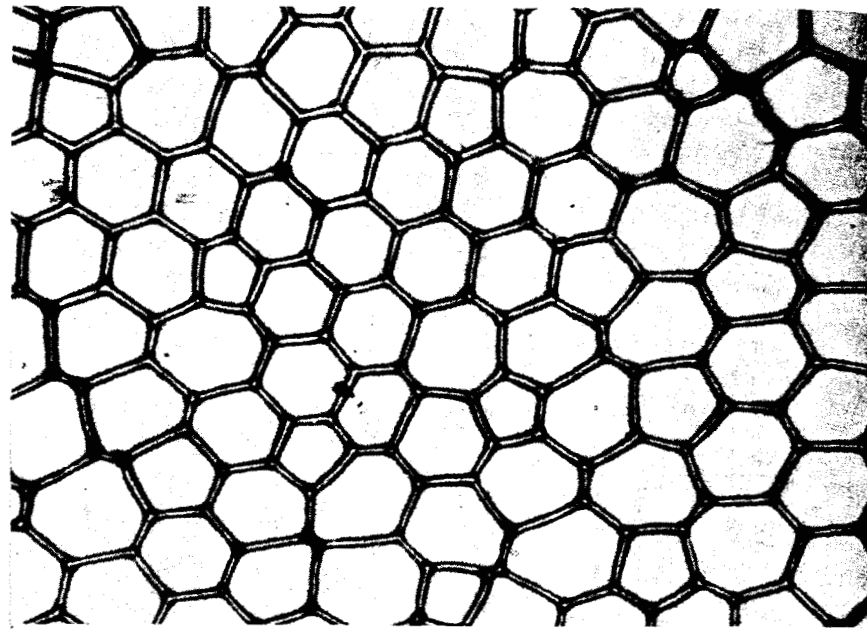


Fig. 4.26 Cellular microstructures. (a) A decanted interface of a cellularly solidified Pb-Sn alloy ($\times 120$) (after J.W. Rutter in *Liquid Metals and Solidification*, American Society for Metals, 1958, p. 243). (b) Longitudinal view of cells in carbon tetrabromide ($\times 100$) (after K.A. Jackson and J.D. Hunt, *Acta Metallurgica* **13** (1965) 1212).



Fig. 4.27 Cellular dendrites in carbon tetrabromide. (After L.R. Morris and W.C. Winegard, *Journal of Crystal Growth* **6** (1969) 61.)

causing recesses to form (c), which in turn trigger the formation of other protrusions (d). Eventually the protrusions develop into long arms or *cells* growing parallel to the direction of heat flow (e). The solute rejected from the solidifying liquid concentrates into the cell walls which solidify at the lowest temperatures. The tips of the cells, however, grow into the hottest liquid and therefore contain the least solute. Even if $X_0 \ll X_{\max}$ (Fig. 4.19) the liquid between the cells may reach the eutectic composition in which case the cell walls will contain a second phase. The interaction between temperature gradient, cell shape and solute segregation is shown in Fig. 4.25. Figure 4.26 shows the appearance of the cellular structure. Note that each cell has virtually the same orientation as its neighbours and together they form a single grain.

Cellular microstructures are only stable for a certain range of temperature gradients. At sufficiently low temperature gradients the cells, or primary arms of solid, are observed to develop secondary arms, and at still lower temperature gradients tertiary arms develop, i.e. dendrites form. Concomitant with

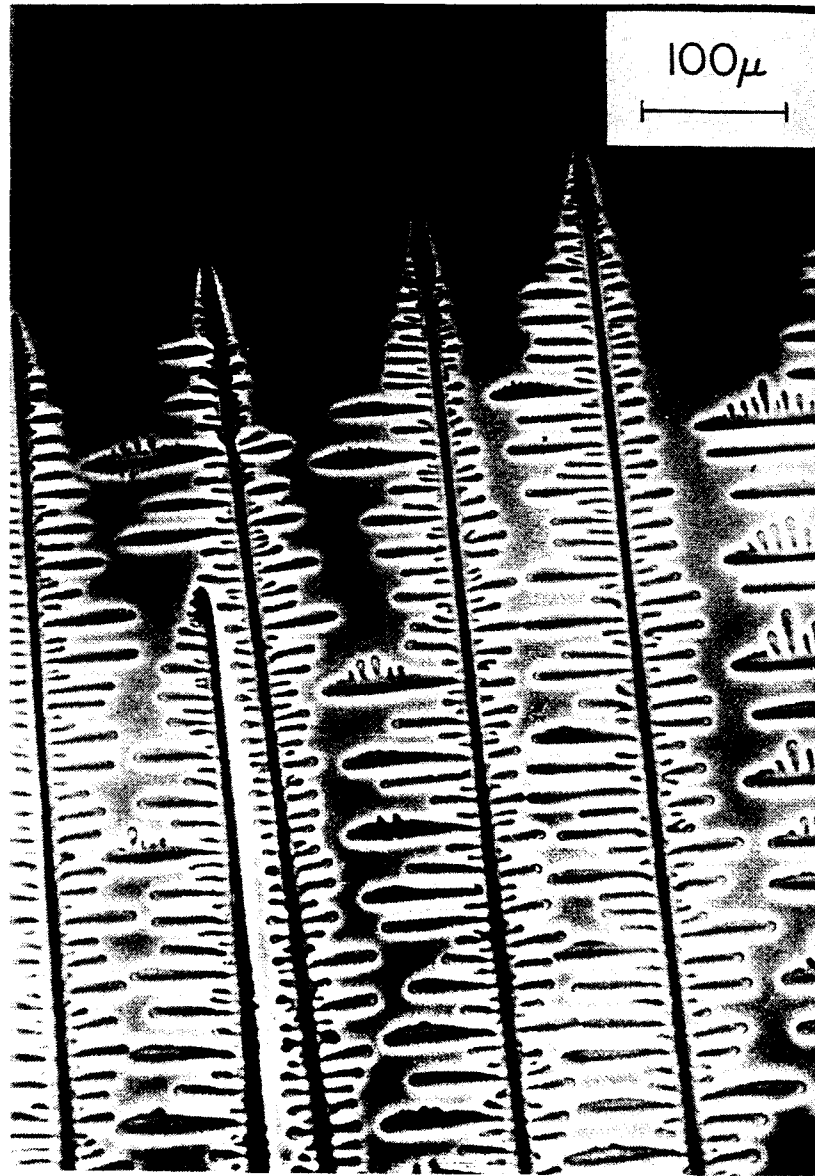


Fig. 4.28 Columnar dendrites in a transparent organic alloy. (After K.A. Jackson in *Solidification*, American Society for Metals, 1971, p. 121.)

this change in morphology there is a change in the direction of the primary arms away from the direction of heat flow into the crystallographically preferred directions such as $\langle 100 \rangle$ for cubic metals. The change in morphology from cells to dendrites can be seen in Figs. 4.26b, 4.27 and 4.28. These pictures have been taken during *in situ* solidification of special transparent

organic compounds using a transmission light microscope⁸. The compounds used have low entropies of melting and solidify in the same way as metals. Alloys have been simulated by suitable combinations of such compounds.

In general the tendency to form dendrites increases as the solidification range increases. Therefore the effectiveness of different solutes can vary widely. For solutes with a very small partition coefficient (k) cellular or dendritic growth can be caused by the addition of a very small fraction of a per cent solute.

The reason for the change from cells to dendrites is not fully understood. However it is probably associated with the creation of constitutional supercooling in the liquid between the cells causing interface instabilities in the transverse direction. Note that for unidirectional solidification there is approximately no temperature gradient perpendicular to the growth direction. The cell or dendrite arm spacing developing is probably that which reduces the constitutional supercooling in the intervening liquid to a very low level. This would be consistent with the observation that cell and dendrite arm spacings both decrease with increasing cooling rate: higher cooling rates allow less time for lateral diffusion of the rejected solute and therefore require smaller cell or dendrite arm spacings to avoid constitutional supercooling.



Fig. 4.29 Al-Cu Al_2 lamellar eutectic normal to the growth direction ($\times 680$). (Courtesy of J. Strid, University of Luleå, Sweden.)



Fig. 4.30 Rod-like eutectic. Al_6Fe rods in Al matrix. Transverse section. Transmission electron micrograph ($\times 70\,000$). (Courtesy of J. Strid, University of Luleå, Sweden.)

Finally it should be noted that although the discussion of alloy solidification has been limited to the case $k < 1$, similar arguments can be advanced for the case of $k > 1$. (See exercise 4.13.)

4.3.2 Eutectic Solidification⁹

In the solidification of a binary eutectic composition two solid phases form cooperatively from the liquid, i.e. $L \rightarrow \alpha + \beta$. Various different types of eutectic solidification are possible and these are usually classified as *normal* and *anomalous*. In normal structures the two phases appear either as alternate lamellae, Fig. 4.29, or as rods of the minor phase embedded in the other phase, Fig. 4.30. During solidification both phases grow simultaneously behind an essentially planar solid/liquid interface. Normal structures occur when both phases have low entropies of fusion. Anomalous structures, on the other hand, occur in systems when one of the solid phases is capable of faceting, i.e. has a high entropy of melting. There are many variants of these structures the most important commercially being the flake structure of Al—Si alloys. This section will only be concerned with normal structures, and deal mainly with lamellar morphologies.

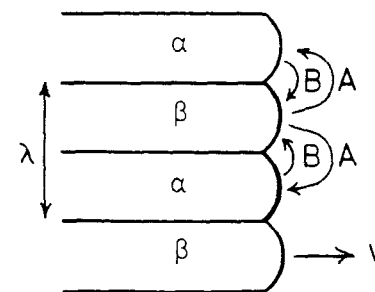


Fig. 4.31 Interdiffusion in the liquid ahead of a eutectic front.

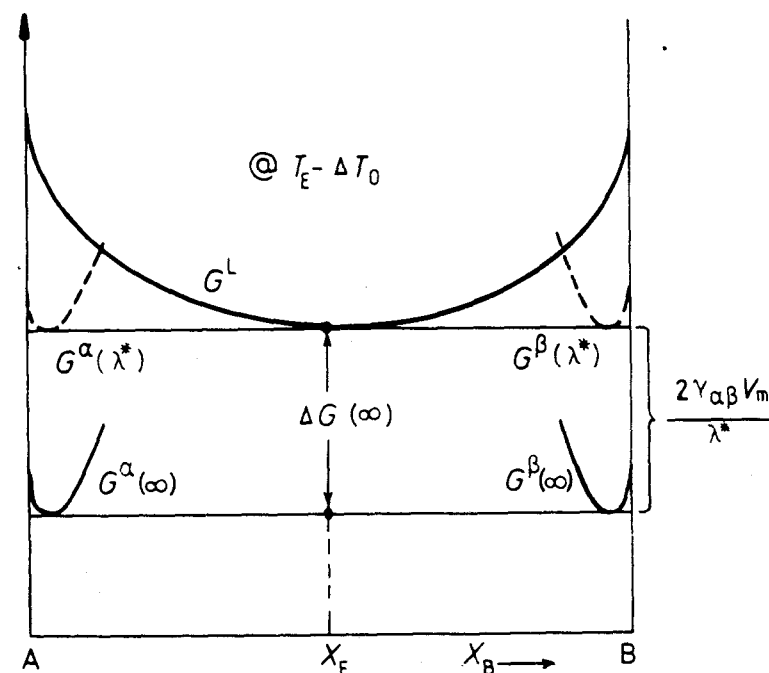


Fig. 4.32 Molar free energy diagram at a temperature ΔT_0 below the eutectic temperature, for the case $\lambda = \lambda^*$.

Growth of Lamellar Eutectics

Figure 4.31 shows how two phases can grow cooperatively behind an essentially planar solidification front. As the A-rich α phase solidifies excess B diffuses a short distance laterally where it is incorporated in the B-rich β phase. Similarly the A atoms rejected ahead of the β diffuse to the tips of the adjacent α lamellae. The rate at which the eutectic grows will depend on how fast this diffusion can occur and this in turn will depend on the interlamellar spacing λ . Thus small interlamellar spacings should lead to rapid growth.

However there is a lower limit to λ determined by the need to supply the α/β interfacial energy, $\gamma_{\alpha\beta}$.

For an interlamellar spacing, λ , there is a total of $(2/\lambda)$ m² of α/β interface per m³ of eutectic. Thus the free energy change associated with the solidification of 1 mol of liquid is given by

$$\Delta G(\lambda) = -\Delta G(\infty) + \frac{2\gamma_{\alpha\beta}V_m}{\lambda} \quad (4.37)$$

where V_m is the molar volume of the eutectic and $\Delta G(\infty)$ is the free energy decrease for very large values of λ . Since solidification will not take place if $\Delta G(\lambda)$ is positive, $\Delta G(\infty)$ must be large enough to compensate for the interfacial energy term, i.e. the eutectic/liquid interface must be undercooled below the equilibrium eutectic temperature T_E , Fig. 4.32. If the total undercooling is ΔT_0 it can be shown that $\Delta G(\infty)$ is then given approximately by

$$\Delta G(\infty) = \frac{\Delta H \cdot \Delta T_0}{T_E} \quad (4.38)$$

where ΔH is an enthalpy term. The minimum possible spacing (λ^*) is obtained by using the relation $\Delta G(\lambda^*) = 0$, whence

$$\lambda^* = \frac{2\gamma_{\alpha\beta}V_mT_E}{\Delta H \cdot \Delta T_0} \quad (4.39)$$

When the eutectic has this spacing the free energy of the liquid and eutectic is the same, i.e. all three phases are in equilibrium. This is because the α/β interface raises the free energies of the α and β from $G^\alpha(\infty)$ and $G^\beta(\infty)$ to $G^\alpha(\lambda^*)$ and $G^\beta(\lambda^*)$ as shown in Fig. 4.32. The cause of the increase is the curvature of the α/L and β/L interfaces arising from the need to balance the interfacial tensions at the $\alpha/\beta/L$ triple point, Fig. 4.31. In general, therefore, the increase will be different for the two phases, but for simple cases it can be shown to be $2\gamma_{\alpha\beta}V_m/\lambda$ for both, Fig. 4.32.

Let us now turn to the mechanism of growth. If solidification is to occur at a finite rate there must be a flux of atoms between the tips of the α and β lamellae and this requires a finite composition difference. For example the concentration of B must be higher ahead of the α phase than ahead of the β phase so that B rejected from the α can diffuse to the tips of the growing β . If $\lambda = \lambda^*$ growth will be infinitely slow because the liquid in contact with both phases has the same composition, X_E in Fig. 4.32. However if the chosen spacing is greater than λ^* less free energy is locked in the interfaces and G^α and G^β are correspondingly reduced, Fig. 4.33a. Under these circumstances the liquid in local equilibrium with the α has a composition $X_B^{L/\alpha}$ which is richer in B than the composition in equilibrium with the β phase $X_B^{L/\beta}$.

If the α/L and β/L interfaces are highly mobile it is reasonable to assume that growth is diffusion controlled in which case the eutectic growth rate (v) should be proportional to the flux of solute through the liquid. This in turn will vary as $D \, dC/dl$ where D is the liquid diffusivity and dC/dl is the

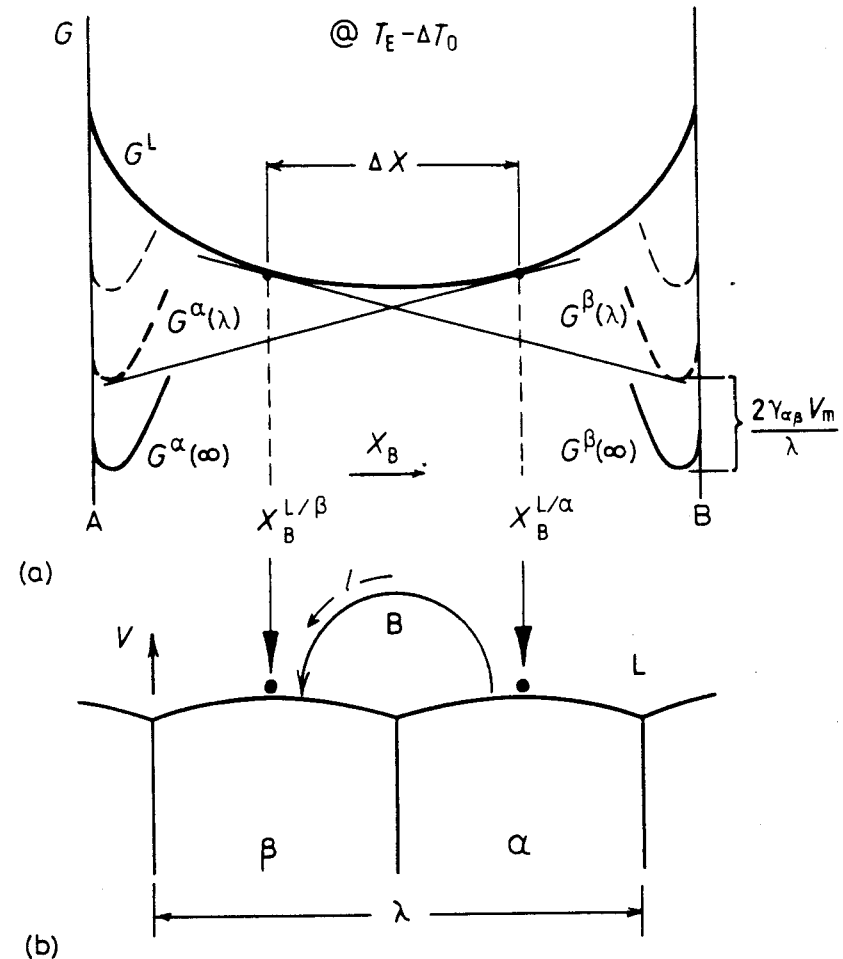


Fig. 4.33 (a) Molar free energy diagram at $(T_E - \Delta T_0)$ for the case $\lambda^* < \lambda < \infty$, showing the composition difference available to drive diffusion through the liquid (ΔX). (b) Model used to calculate the growth rate.

concentration gradient driving the diffusion. (l is measured along the direction of diffusion as shown in Fig. 4.33b. In practice dC/dl will not have a single value but will vary from place to place within the diffusion zone.) dC/dl should be roughly proportional to the maximum composition difference ($X_B^{L/\alpha} - X_B^{L/\beta}$) and inversely proportional to the effective diffusion distance, which, in turn, will be linearly related to the interlamellar spacing (λ). Thus we can write

$$v = k_1 D \frac{\Delta X}{\lambda} \quad (4.40)$$

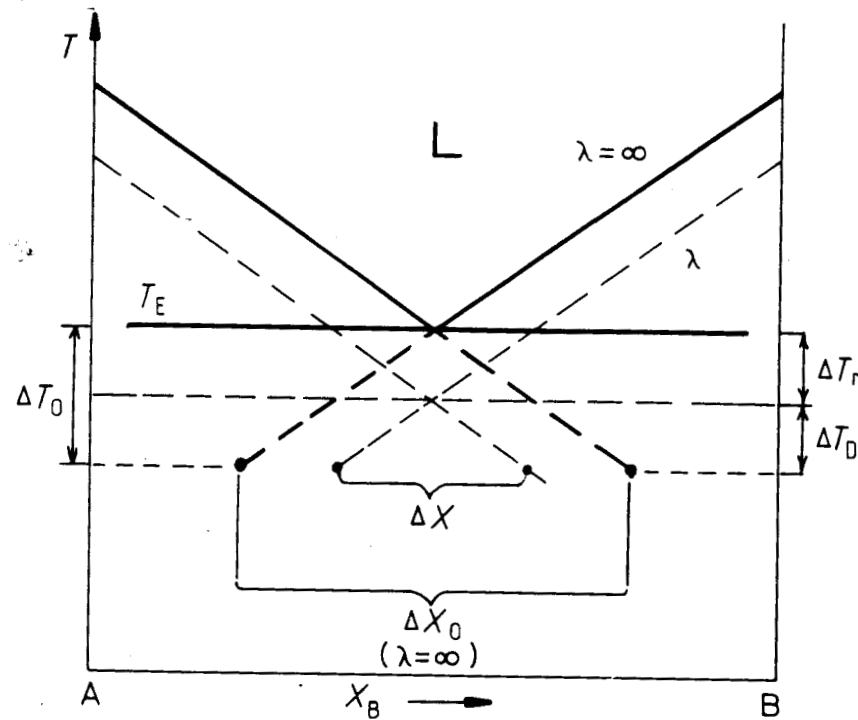


Fig. 4.34 Eutectic phase diagram showing the relationship between ΔX and ΔX_0 (exaggerated for clarity).

where k_1 is a proportionality constant and $\Delta X = X_B^{L/\alpha} - X_B^{L/\beta}$ as given in Fig. 4.33.

ΔX will itself depend on λ for when $\lambda = \lambda^*$, $\Delta X = 0$, and as λ increases ΔX will tend to a maximum value, ΔX_0 , say. Therefore it is reasonable to write

$$\Delta X = \Delta X_0 \left(1 - \frac{\lambda^*}{\lambda} \right) \quad (4.41)$$

The magnitude of ΔX_0 can be obtained by extrapolating the equilibrium liquidus lines of the phase diagram ($\lambda = \infty$) as shown in Fig. 4.34. For small undercoolings

$$\Delta X_0 \propto \Delta T_0 \quad (4.42)$$

The dashed lines in Fig. 4.34 are the liquidus lines for $\lambda^* < \lambda < \infty$. ΔX is simply given by the extrapolation of these lines as shown. Combining the above equations gives

$$v = k_2 D \Delta T_0 \frac{1}{\lambda} \left(1 - \frac{\lambda^*}{\lambda} \right) \quad (4.43)$$

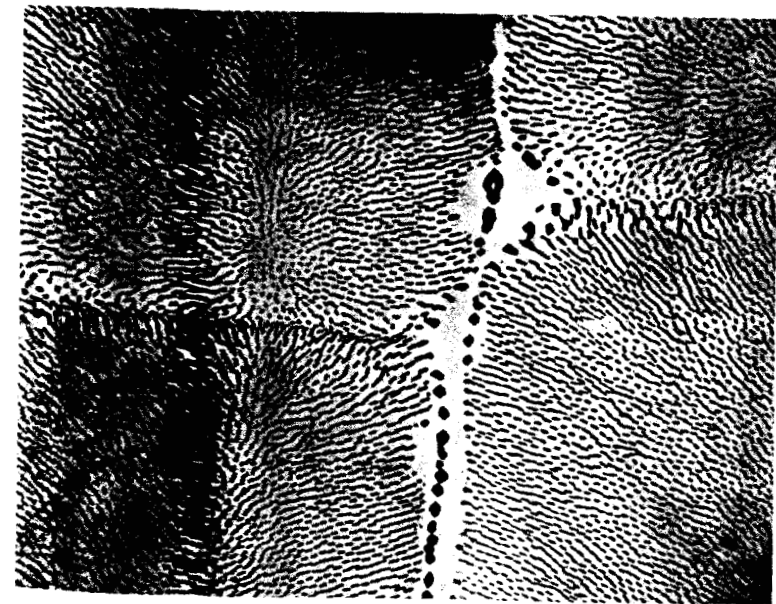


Fig. 4.35 (a) Cellular eutectic solidification front in a transparent organic alloy. (After J.D. Hunt and K.A. Jackson, *Transactions of the Metallurgical Society of AIME* 236 (1966) 843.) (b) Transverse section through the cellular structure of an Al-Al₆Fe rod eutectic (3500). (Courtesy of J. Strid, University of Luleå, Sweden.)

where k_2 is another proportionality constant. This equation shows that by varying the interface undercooling (ΔT_0) it is possible to vary the growth rate (v) and spacing (λ) independently. It is therefore impossible to predict the spacing that will be observed for a given growth rate. However, controlled growth experiments show that a specific value of λ is always associated with a given growth rate. Examination of Equation 4.43 shows that when $\lambda = 2\lambda^*$, the growth rate is a maximum for a given undercooling, or, alternatively, the required undercooling is a minimum for a given growth rate. If it is assumed that growth occurs under these optimum conditions the observed spacing $\lambda_0 = 2\lambda^*$ and the observed growth rate is given by

$$v_0 = k_2 D \Delta T_0 / 2\lambda^*$$

However, from Equation 4.39, it is seen that $\Delta T_0 \propto 1/\lambda^*$ so that the following relationships are predicted:

$$v_0 \lambda_0^2 = k_3 \text{ (constant)} \quad (4.44)$$

and

$$\frac{v_0}{(\Delta T_0)^2} = k_4 \text{ (constant)} \quad (4.45)$$

There is in fact no physical basis for choosing $\lambda = 2\lambda^*$ and similar expressions can also be obtained using other assumptions concerning the spacing. Equations 4.44 and 4.45 are often found to be obeyed experimentally. For example measurements on the lamellar eutectic in the Pb-Sn system¹⁰ show that $k_3 \sim 33 \mu\text{m}^3 \text{s}^{-1}$ and $k_4 \sim 1 \mu\text{m s}^{-1} \text{K}^{-2}$. Therefore for a solidification rate of $1 \mu\text{m s}^{-1}$, $\lambda_0 \sim 5 \mu\text{m}$ and $\Delta T_0 \sim 1 \text{K}$.

The total undercooling at the eutectic front (ΔT_0) has two contributions, i.e.

$$\Delta T_0 = \Delta T_r + \Delta T_D \quad (4.46)$$

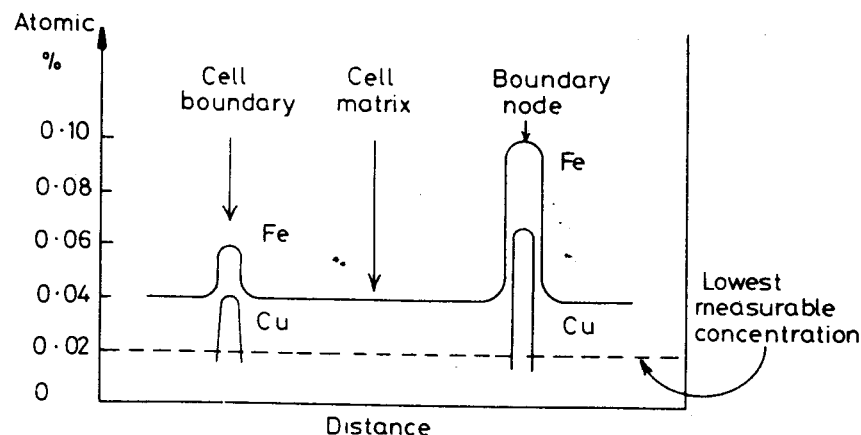


Fig. 4.36 Composition profiles across the cells in Fig. 4.35b.

ΔT_r is the undercooling required to overcome the interfacial curvature effects and ΔT_D is the undercooling required to give a sufficient composition difference to drive the diffusion. (Strictly speaking a ΔT_i term should also be added since a driving force is required to move the atoms across the interfaces, but this is negligible for high mobility interfaces.) A better theoretical treatment of eutectic solidification should take into account the fact that the composition of the liquid in contact with the interface and therefore ΔT_D vary continuously from the middle of the α to the middle of the β lamellae. Since the interface is essentially isothermal (ΔT_0 constant) the variation of ΔT_D must be compensated by a variation in ΔT_r , i.e. the interface curvature will change across the interface¹¹.

A planar eutectic front is not always stable. If for example the binary eutectic alloy contains impurities, or if other alloying elements are present, the interface tends to break up to form a cellular morphology. The solidification direction thus changes as the cell walls are approached and the lamellar or rod structure fans out and may even change to an irregular structure, Fig. 4.35. The impurity elements diffuse laterally and concentrate at the cell walls. In the case of the Al₆Fe—Al rod-like eutectic shown in Fig. 4.35 the impurity causing the cellular structure is mainly copper. Figure 4.36 shows how the concentration of copper and iron in the aluminium matrix increases in the cell walls and boundary nodes.

Cell formation in eutectic structures is analogous to that in single-phase solidification, and under controlled conditions it is possible to stabilize a planar interface by solidifying in a sufficiently high temperature gradient.

4.3.3 Off-Eutectic Alloys

When the bulk alloy composition (X_0) deviates from the equilibrium eutectic composition (X_E) as shown in Fig. 4.37 solidification usually begins close to T_1 with the formation of primary (α) dendrites. As the dendrites thicken solute is rejected into the remaining liquid until its composition reaches X_E and the eutectic solidifies. Under steady-state unidirectional solidification conditions in the presence of a shallow temperature gradient the solidification front could appear as in Fig. 4.37b. The tips of the dendrites are close to T_1 and the eutectic front, most probably cellular, close to T_E . Similar behaviour is found during the solidification of castings and ingots. In the absence of solid-state diffusion the centres of the dendrites, which solidified close to T_1 , will contain less solute than the outer layers that solidify at progressively lower temperatures. This leads to what is known as *coring* in the final microstructure, Fig. 4.38. The eutectic does not always solidify as a two-phase mixture. When the volume fraction of one of the phases in the microstructure is very small it can form a so-called *divorced eutectic*. The minor phase then often appears as isolated islands and the other phase forms by the thickening of the dendrites.

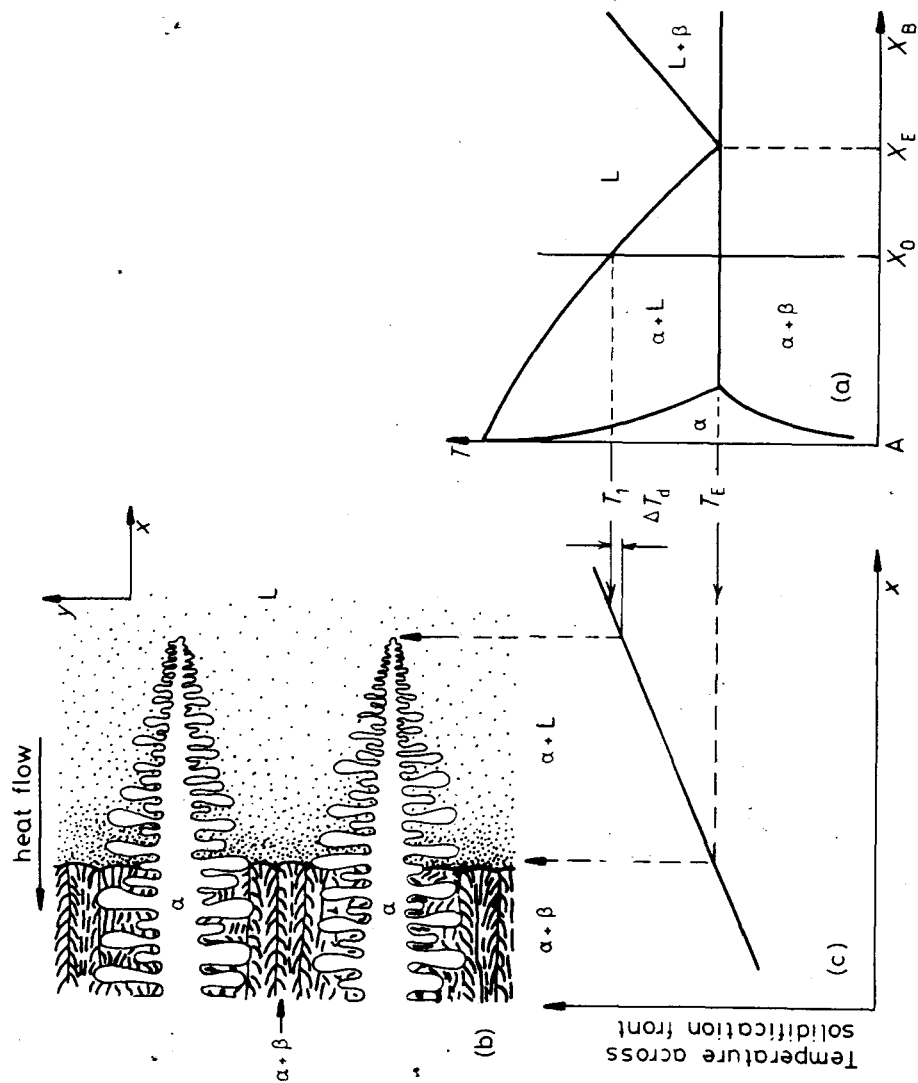


Fig. 4.37. Solidification of an off-eutectic alloy in a temperature gradient. (a) Alloy composition in relation to the phase diagram. (b) Schematic solidification front. (c) Temperature variation across solidification front (isotherms normal to x).

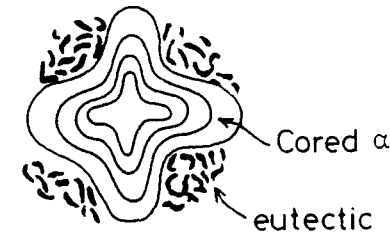


Fig. 4.38 Transverse section through a dendrite in Fig. 4.37.

Under controlled solidification conditions, e.g. in unidirectional solidification experiments, it is possible to solidify an off-eutectic alloy without permitting the formation of the primary dendritic phase. If the temperature gradient in the liquid is raised above a critical level the dendrite tips are overgrown by the eutectic and the alloy solidifies as 100% 'eutectic' with an overall composition X_0 instead of X_E . A similar change can be brought about if the growth rate is raised above a critical level. In both cases the reason for the disappearance of the primary dendrites is that for a given growth velocity the eutectic is able to grow at a higher temperature than the dendrite tips¹². This phenomenon is of special interest in the production of *in situ* composite materials because the volume fraction of the two phases in the composite can be controlled by the choice of X_0 ¹³.

4.3.4 Peritectic Solidification¹⁴

A typical phase diagram containing a peritectic reaction, i.e. $L + \alpha \rightarrow \beta$, is shown in Fig. 4.39a. During equilibrium solidification solid α with composition 'a' and liquid with composition 'c' react at the peritectic temperature T_p to give solid β of composition 'b'. However, the transformation rarely goes to completion in practice.

Consider for example the solidification of an alloy X_0 at a finite velocity in a shallow temperature gradient, Fig. 4.39b and c. As the temperature decreases the first phase to appear is α with the composition $\sim kX_0$ at a temperature close to T_1 . α grows dendritically with successive layers solidifying at compositions determined by the local temperature and the α solidus. If diffusion in the dendrites is slow the liquid will eventually reach point c in Fig. 4.39a and on further cooling it reacts with the α to produce a layer of β . However, the α dendrites are then often effectively isolated from further reaction and are retained to lower temperatures. Meanwhile the β phase continues to precipitate from the liquid at compositions which follow the line bd. Again if there is no diffusion in the solid the liquid will finally reach point e and solidify as a $\beta + \gamma$ eutectic. The final solidified microstructure will then consist of cored α dendrites surrounded by a layer of β and islands of $\beta + \gamma$ eutectic, or divorced eutectic.

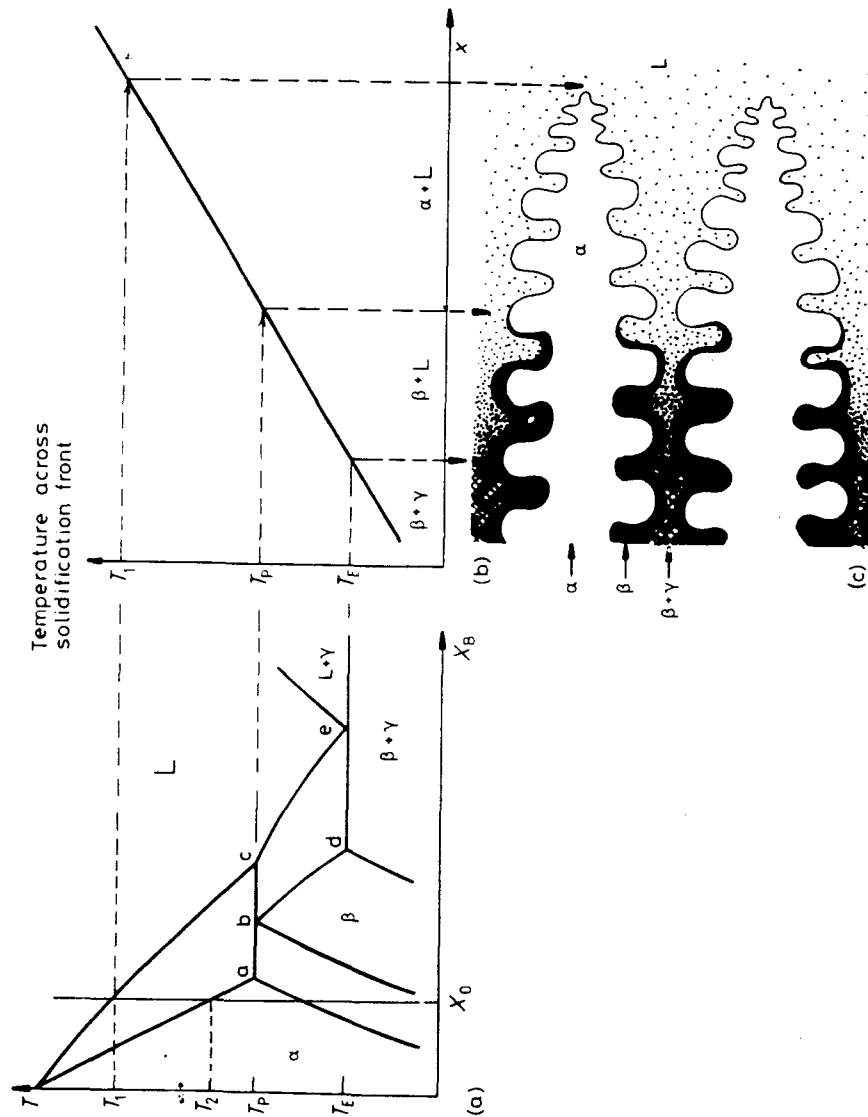


Fig. 4.39 Peritectic solidification in a temperature gradient.

If alloy X_0 were directionally solidified at increasing values of (T'_L/v) the temperature of the dendrite tips would progressively fall from T_1 towards T_2 (Fig. 4.39a) while the temperature at which the last liquid solidifies would increase towards T_2 . Finally, solidification would take place behind a planar front at a temperature T_2 , as discussed earlier in Section 4.3.1.

Planar-front solidification can also be obtained for alloys beyond 'a' in Fig. 4.39, provided a sufficiently high temperature-gradient/velocity ratio is used. Alloys between a and b then solidify with a 'eutectic-like' $\alpha + \beta$ structure. (The structure is better described as *composite* to avoid confusion concerning the term eutectic). Between b and d single-phase β forms, and beyond d $\beta + \gamma$ eutectic-like structures can be formed.

The Fe—C phase diagram also contains a peritectic, Fig. 4.53a. However due to the high diffusivity of carbon at these high temperatures the peritectic reaction is very rapid and is able to convert all of the primary (δ) dendrites into the more stable austenite.

4.4 Solidification of Ingots and Castings

This section is concerned with technological applications of the theory of solidification, as developed earlier. Two of the most important applications are casting and weld solidification and we shall first consider these. In modern constructions there is a tendency towards the use of stronger, heavier sections welded with higher energy techniques and faster speeds. It is thus important for the physical metallurgist to consider the effect of the various solidification parameters on the microstructure and properties of fusion welds. This will then be followed by some selected case studies of as-solidified or as-welded engineering alloys and weld metals.

Most engineering alloys begin by being poured or cast into a fireproof container or *mould*. If the as-cast pieces are permitted to retain their shape afterwards, or are reshaped by machining, they are called *castings*. If they are later to be worked, e.g. by rolling, extrusion or forging, the pieces are called *ingots*, or *blanks* if they are relatively small. In either case the principles of solidification, and the requirements for high density and strength are the same. The moulds used in casting are often made of a material that can be remoulded or discarded after a casting series, such as sand. In the case of long casting series or ingot casting, however, the mould is of a more permanent material such as cast iron. The technological aspects of pouring and casting will not be dealt with here, but we shall confine our discussion simply to the mechanics of solidification of metals in a mould.

4.4.1 Ingot Structure

Generally speaking three different zones can be distinguished in solidified alloy ingots, Fig. 4.40. These are (i) an outer *chill zone* of equiaxed crystals, (ii) a *columnar zone* of elongated or column-like grains, and (iii) a central *equiaxed zone*.

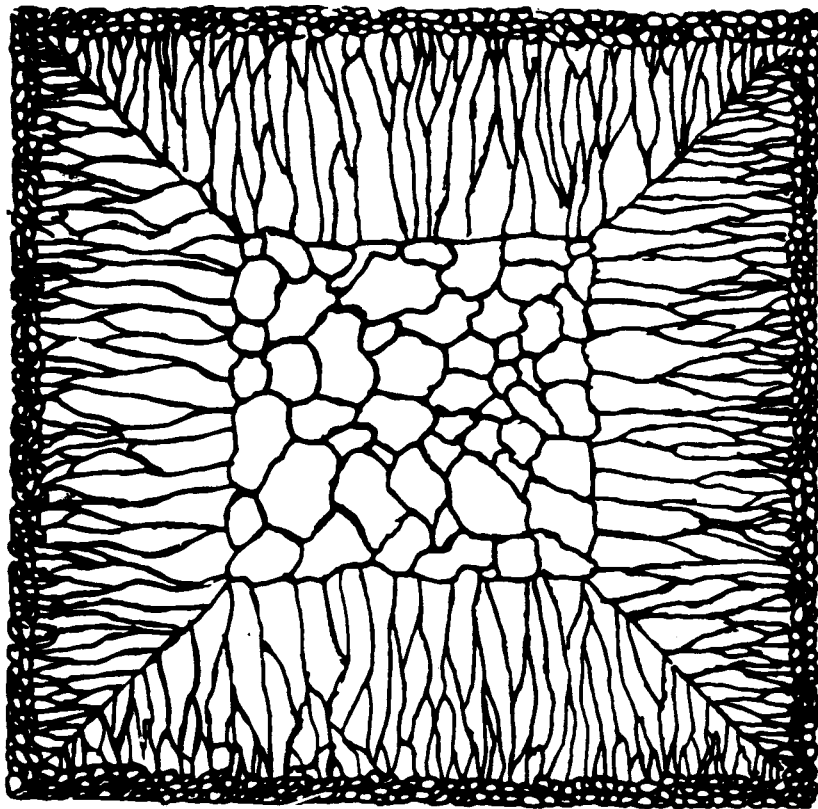


Fig. 4.40 Schematic cast grain structure. (After M.C. Flemings, *Solidification Processing*, McGraw-Hill, New York, 1974.)

Chill Zone

During pouring the liquid in contact with the cold mould wall is rapidly cooled below the liquidus temperature. Many solid nuclei then form on the mould wall and begin to grow into the liquid, Fig. 4.41. As the mould wall warms up it is possible for many of these solidified crystals to break away from the wall under the influence of the turbulent melt. If the pouring temperature is low the whole of the liquid will be rapidly cooled below the liquidus temperature and the crystals swept into the melt may be able to continue to grow. This is known as 'big-bang' nucleation since the liquid is immediately filled with a myriad of crystals. This produces an entirely equiaxed ingot structure, i.e. no columnar zone forms. If the pouring temperature is high, on the other hand, the liquid in the centre of the ingot will remain above the liquidus temperature for a long time and consequently the majority of crystals soon remelt after breaking away from the mould wall. Only those crystals remaining close to the wall will be able to grow to form the chill zone.

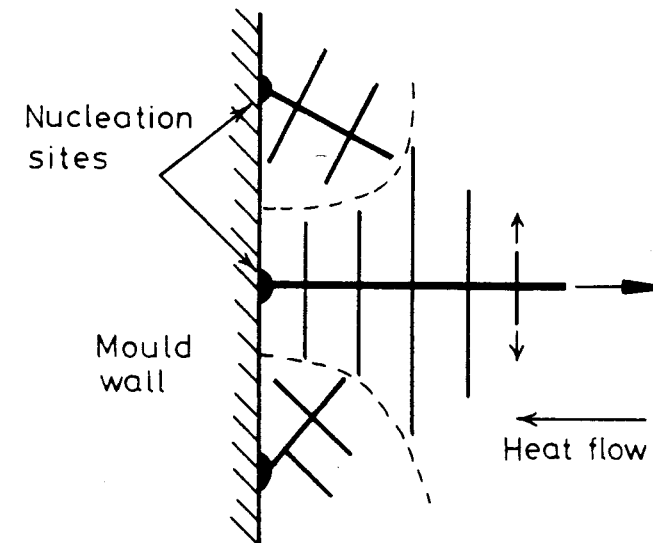


Fig. 4.41 Competitive growth soon after pouring. Dendrites with primary arms normal to the mould wall, i.e. parallel to the maximum temperature gradient, outgrow less favourably oriented neighbours.

Columnar Zone

Very soon after pouring the temperature gradient at the mould walls decreases and the crystals in the chill zone grow dendritically in certain crystallographic directions, e.g. $\langle 100 \rangle$ in the case of cubic metals. Those crystals with a $\langle 100 \rangle$ direction close to the direction of heat flow, i.e. perpendicular to the mould walls, grow fastest and are able to outgrow less favourably oriented neighbours, Fig. 4.42. This leads to the formation of the columnar grains all

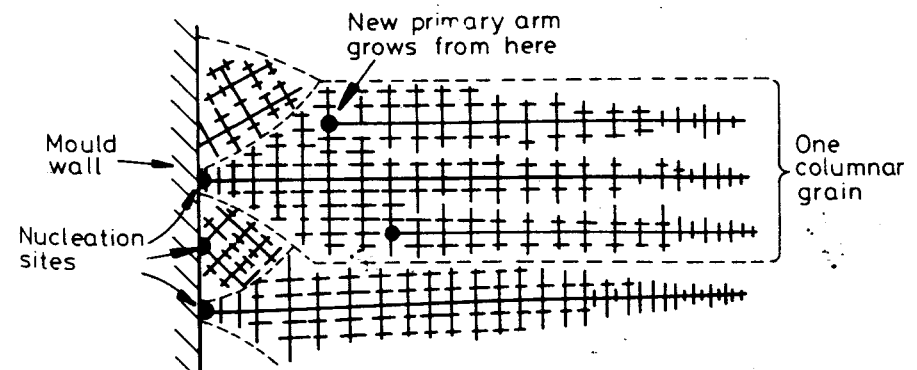


Fig. 4.42 Favourably oriented dendrites develop into columnar grains. Each columnar grain originates from the same heterogeneous nucleation site, but can contain many primary dendrite arms.

with $\langle 100 \rangle$ almost parallel to the column axis. Note that each columnar crystal contains many primary dendrite arms. As the diameter of these grains increases additional primary dendrite arms appear by a mechanism in which some tertiary arms grow ahead of their neighbours as shown in the figure.

The volume fraction of the melt solidified increases with increasing distance behind the tips of the dendrites and, when the structure is mainly single-phase, the secondary and tertiary arms of adjacent dendrites can link up to form walls of solid containing many primary dendrite arms. The region between the tips of the dendrites and the point where the last drop of liquid is solidifying is known as the *mushy* or *pasty* zone. The length of this zone depends on the temperature gradient and the non-equilibrium freezing range of the alloy. In general it is found that the secondary arms become coarser with distance behind the primary dendrite tips. This effect can be seen in Fig. 4.28. The primary and secondary dendrite arm spacing is also often found to increase with increasing distance from the mould wall. This is simply due to a corresponding decrease in the cooling rate with time after pouring.

Equiaxed Zone

The equiaxed zone consists of equiaxed grains randomly oriented in the centre of the ingot. An important origin of these grains is thought to be melted-off dendrite side-arms. It can be seen from Fig. 4.28 that the side-arms are narrowest at their roots. Therefore, if the temperature around the dendrite increases after it has formed, it will begin to melt and may become detached from the main stem. Provided the temperature falls again before the arm completely disappears it can then act as a 'seed' for a new dendrite. An effective source of suitable temperature pulses is provided by the turbulent convection currents in the liquid brought about by the temperature differences across the remaining melt. Convection currents also provide a means of carrying the melted-off arms away to where they can develop uninhibited into equiaxed dendrites. If convection is reduced fewer seed crystals are created causing a larger final grain size and a greater preponderance of columnar grains. Convection also plays a dominant role in the formation of the outer chill zone. The mechanism whereby crystals are melted away from the mould walls is thought to be similar to the detachment of side-arms¹⁵ and when convection is absent no chill zone is observed.

Shrinkage Effects

Most metals shrink on solidification and this has important consequences for the final ingot structure. In alloys with a narrow freezing range the mushy zone is also narrow and as the outer shell of solid thickens the level of the remaining liquid continually decreases until finally when solidification is complete the ingot contains a deep central cavity or *pipe*.

In alloys with a wide freezing range the mushy zone can occupy the whole of the ingot. In this case no central pipe is formed. Instead the liquid level gradually falls across the width of the ingot as liquid flows down to compen-

sate for the shrinkage of the dendrites. However, as the interdendritic channels close up this liquid flow is inhibited so that the last pools of liquid to solidify leave small voids or pores.

4.4.2 Segregation in Ingots and Castings

Two types of segregation can be distinguished in solidified structures. There is *macrosegregation*, i.e. composition changes over distances comparable to the size of the specimen, and there is *microsegregation* that occurs on the scale of the secondary dendrite arm spacing.

It has already been shown that large differences in composition can arise across the dendrites due to coring and the formation of non-equilibrium phases in the last solidifying drops of liquid. Experimentally it is found that while cooling rate affects the spacing of the dendrites it does not substantially alter the amplitude of the solute concentration profiles provided the dendrite morphology does not change and that diffusion in the solid is negligible. This result often applies to quite a wide range of practical cooling rates.

There are four important factors that can lead to macrosegregation in ingots. These are: (i) shrinkage due to solidification and thermal contraction; (ii) density differences in the interdendritic liquid; (iii) density differences between the solid and liquid; and (iv) convection currents driven by temperature-induced density differences in the liquid. All of these factors can induce macrosegregation by causing mass flow over large distances during solidification.

Shrinkage effects can give rise to what is known as *inverse segregation*. As the columnar dendrites thicken solute-rich liquid (assuming $k < 1$) must flow back between the dendrites to compensate for shrinkage and this raises the solute content of the outer parts of the ingot relative to the centre. The effect is particularly marked in alloys with a wide freezing range, e.g. Al-Cu and Cu-Sn alloys.

Interdendritic liquid flow can also be induced by gravity effects. For example during the solidification of Al-Cu alloys the copper rejected into the liquid raises its density and causes it to sink. The effect can be reinforced by convection currents driven by temperature differences in the ingot.

Gravity effects can also be observed when equiaxed crystals are forming. The solid is usually denser than the liquid and sinks carrying with it less solute than the bulk composition (assuming $k < 1$). This can, therefore, lead to a region of negative segregation near the bottom of the ingot.

The combination of all the above effects can lead to complex patterns of macrosegregation. Fig. 4.43 for example illustrates the segregation patterns found in large steel ingots¹⁶.

In general segregation is undesirable as it has marked deleterious effects on mechanical properties. The effects of microsegregation can be mitigated by subsequent homogenization heat treatment, but diffusion in the solid is far too

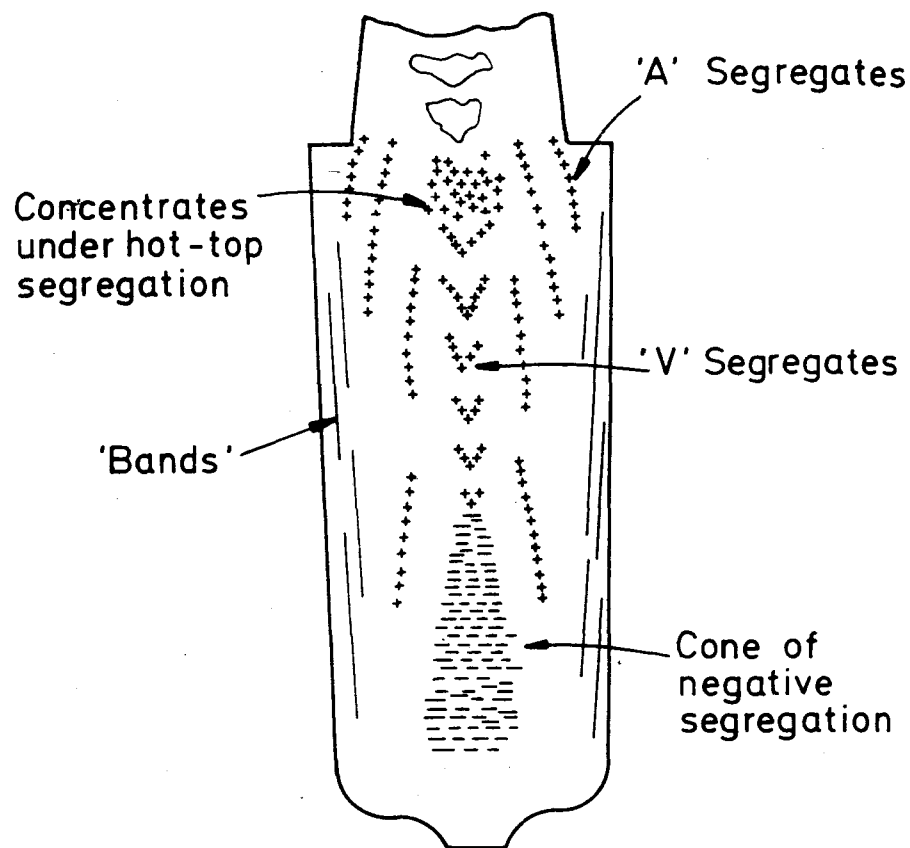


Fig. 4.43 Segregation pattern in a large killed steel ingot. + positive, - negative segregation. (After M.C. Flemings, *Scandinavian Journal of Metallurgy* 5 (1976) 1.)

slow to be able to remove macrosegregation which can only be combated by good control of the solidification process.

4.4.3 Continuous Casting

A number of industrial processes are nowadays employed in which casting is essentially a dynamic rather than a static process. In these cases, the molten metal is poured continuously into a water-cooled mould (e.g. copper) from which the solidified metal is continuously withdrawn in plate or rod form. This process is illustrated schematically in Fig. 4.44.

It is seen that the speed of withdrawal is such that the solid-liquid interface is maintained in the shape and position illustrated. Ideally, the flow behaviour of the liquid should be vertically downwards, and if flow is maintained in this way the final composition across the ingot will be kept uniform. In practice,

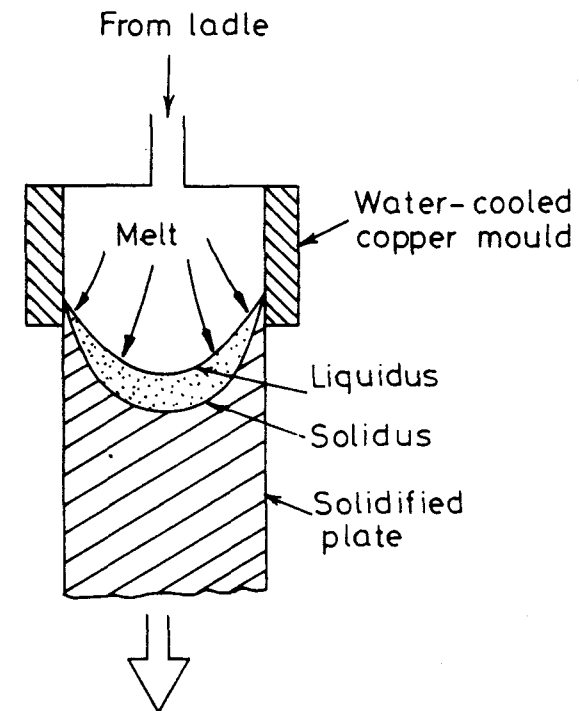


Fig. 4.44 Schematic illustration of a continuous casting process.

hydrodynamic effects do not allow this simple type of flow and there is a tendency for the flow lines to fan outwards (as shown by the arrows) producing negative segregation near the centre. Solidification follows the maximum temperature gradient in the melt as given by the normals to the isotherms. In certain respects weld solidification has much in common with continuous casting in that it is also a dynamic process. As illustrated in Fig. 4.45 the main difference is of course that in continuous casting the heat source (as defined by the mould) does not move, whereas in welding the heat source (the electrode) is moving. We shall now consider the latter case in more detail, but it will be found that certain conclusions concerning weld solidification behaviour can well be applied to both processes.

Heat Flow in Welding and Continuous Casting

As discussed earlier, there are many factors concerning heat distribution at the melt zone and the dynamics of the process, which are essentially fairly similar in both continuous casting and welding. As an example we shall first consider the welding process and then discuss how the results may be applied to continuous casting. In contrast to continuous casting, weld solidification involves a 'mould' that has approximately the same composition as the melt.

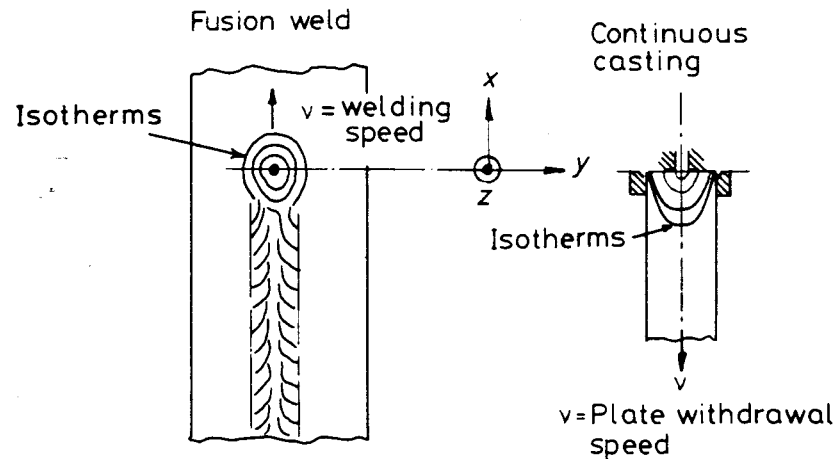


Fig. 4.45 Illustrating the essential equivalence of isotherms about the heat sources in fusion welding and continuous casting.

The most important variables in weld solidification or continuous casting are thus:

1. The rate of heat input, q (determined by type of weld process, weld size, etc.); in terms of continuous casting q is represented effectively by the volume and temperature of the melt.
2. Speed of arc movement along join, v ; in continuous casting, v is the velocity of plate withdrawal.
3. Thermal conductivity of the metal being welded or cast, K_s .
4. Thickness of plate being welded or cast, t .

In the case of welding, assuming that the arc moves along the x coordinate, the resulting heat distribution in a three-dimensional solid plate is given by the solution to the heat flow equation¹⁹:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 2K_s v \frac{\partial T}{\partial (x - vt)} \quad (4.47)$$

where x, y, z are defined in Fig. 4.45 and t is time.

Solving this equation gives the temperature distribution about the moving heat source in the form of isotherms in the solid metal, in which the distance between the isotherms in a given direction (x, y, z) is approximately given by:

$$\lambda_{(x,y,z)} \propto \frac{q}{K_s v t} \quad (4.48)$$

Gray *et al.* (1975) have solved Equation 4.47 and plotted isotherms for a number of different materials and welding speeds and some of their results are summarized in Fig. 4.46.

Assuming a similar isotherm distribution in the melt, it is likely that the

parameters K_s, v, t and q will largely determine solidification morphology, in that dendrites always try to grow in directions as near normal to isotherms as their crystallography allows. It is seen in the above figure, for example, that holding q, v and t constant (Fig. 4.46a), the distance between isotherms, λ , increases substantially as a function of the change of heat conductivity, K_s , of the different materials: aluminium, carbon (ferritic) steel, and austenitic steel. If the plate thickness is increased (Fig. 4.46b), or the weld speed is higher (Fig. 4.46c), λ also decreases proportionally.

These results can also be applied to *continuous casting*, in that the isotherm distributions shown in Fig. 4.46 are affected in a similar way by the conductivity of the solidifying metal and its speed of withdrawal from the mould. This means, for example, that the depth of the liquid pool in continuous casting is much greater for steel than for aluminium alloys under comparable condi-

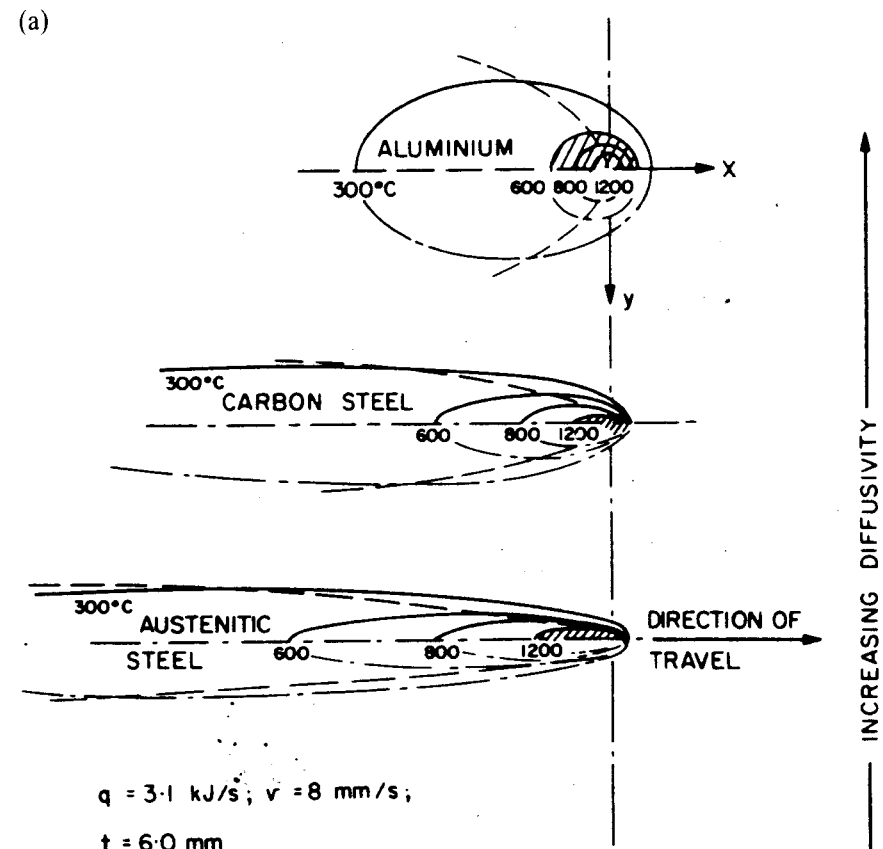


Fig. 4.46 Effect of various parameters in Equation 5.36 on the isotherm distribution at a point heat source. (a) Effect of changes in thermal conductivity, K_s . (b) Effect of changes in plate thickness, t . (c) Effect of changes in movement of heat source, v . (After T.G. Gray, J. Spence and T.H. North, *Rational Welding Design*, Newnes-Butterworth, London, 1975.)

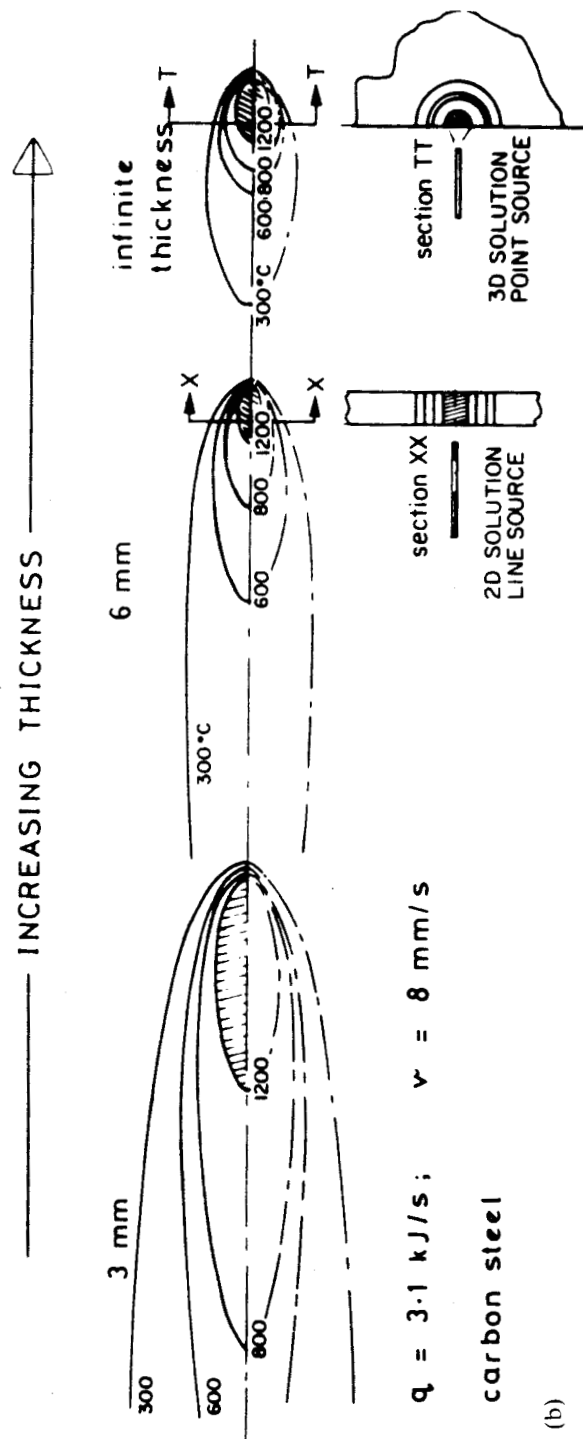
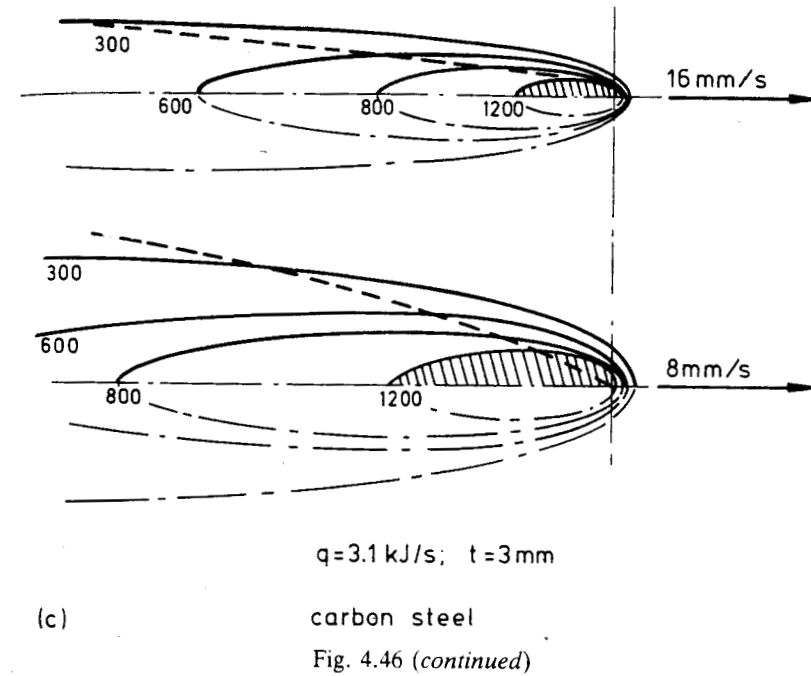


Fig. 4.46 (continued)



tions. This implies that in practice the maximum casting speed and billet cross-section are less for steel than for aluminium or copper. Another practical difficulty resulting from a large depth of liquid is that the billet can not be cut until it reaches a point well beyond the solidus line (see Fig. 4.44), which requires in fact a very tall installation for high speed casting.

4.5 Solidification of Fusion Welds

Contact between the weld melt and the base metal will initially cause melting back of the base material and dilution of the filler metal as illustrated in Fig. 4.47. The amount of dilution involved is not insignificant. Jesseman (1975)¹⁷ reports for example that in microalloyed steel welds, the weld metal may contain 50–70% of the amount of Nb, Ti or V as analysed in the base material. The effect of dilution is in fact threefold, and affects the weld metal

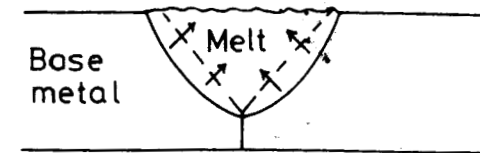


Fig. 4.47 Illustrating the effect of dilution. In high-energy welds, the weld metal typically exhibits 50–70% of the analysis of microalloying elements of the base metal through dilution.

as follows:

1. The composition of the melt is changed.
2. The surface oxide layer of the base metal is removed (also into the melt).
3. It cools down the melt.

Depending upon the type of material being welded as well as plate thickness, the base metal behaves as a very efficient heat sink, and already at $T \approx T_c$ solidification nuclei form at the oxide-free surface of the melted-back base metal. Since the melt has approximately the same composition as the base metal, 'wetting' of the base metal is very efficient and $\theta \approx 0$ (see Fig. 4.7). This implies in turn that there is almost no nucleation barrier to solidification and hence very little undercooling occurs. Solidification is thus predicted to occur epitaxially, i.e. nuclei will have the same lattice structure and orientation as the grains at the solid-liquid surface of the base metal, and this is what is observed in practice.

Since the temperature of the melt beneath the arc is so high and the base material is such an efficient heat sink there is initially a steep temperature gradient in the liquid and consequently the degree of constitutional supercooling is low. The actual thermal gradient is of course dependent upon the welding process and the plate thickness (Equation 4.48). For example TIG welding of thin plates will give steeper thermal gradients than submerged arc

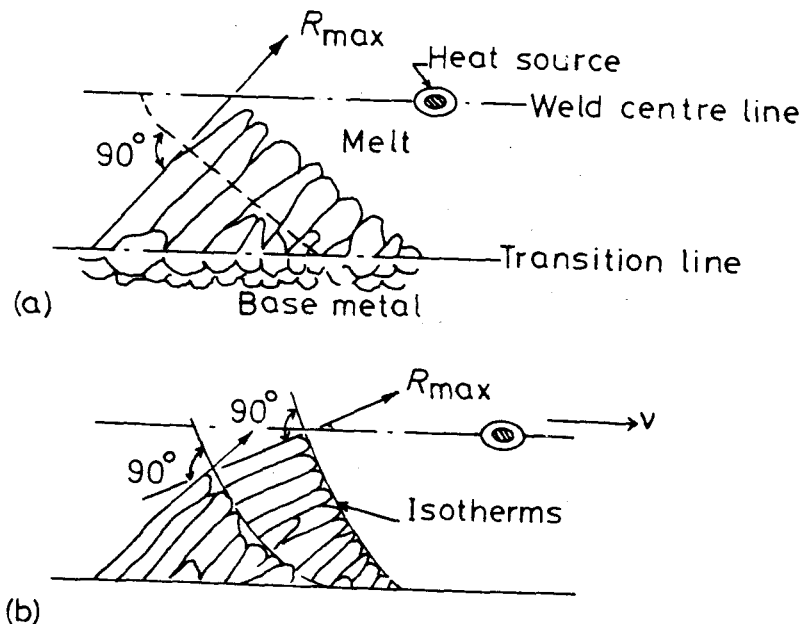


Fig. 4.48 Illustrating the growth of columnar crystals in the weld, and how growth continues to occur approximately normal to the isotherms.

welding of thick plates, the latter process having the higher heat input. Since certain grains at the base metal are better oriented than others for $\langle 100 \rangle$ growth with respect to the isotherms of the melt, these quickly predominate and widen at the expense of the others. However, the general coarseness of the microstructure is largely determined at this stage by the grain size of the base metal. Unfortunately, the base metal at the transition zone receives the most severe thermal cycle and after high-energy welding in particular the grains in this zone tend to grow and become relatively coarse. The weld microstructure is thus inherently coarse grained.

Welding is essentially a dynamic process in which the heat source is continuously moving. This means that the maximum temperature gradients are constantly changing direction as the heat source moves away. The growing columnar crystals are thus faced with the necessity of trying to follow the maximum temperature gradients while still maintaining their preferred $\langle 100 \rangle$ growth direction. This often results in sudden changes in growth direction, as illustrated in Fig. 4.48a and b.

Few of the grains originating at the base metal survive to reach the weld centre line. The mechanism by which sudden changes in $\langle 100 \rangle$ growth direction are brought about is not fully understood. One suggestion is that renucleation occurs by the help of dendritic fragments which have broken away from the growing interface due to turbulence in the weld pool, or simply from melted-off dendrite arms.

Influence of Welding Speed

An important effect of increasing the welding speed is that the shape of the weld pool changes from an elliptical shape to a narrower, pear shape (see.

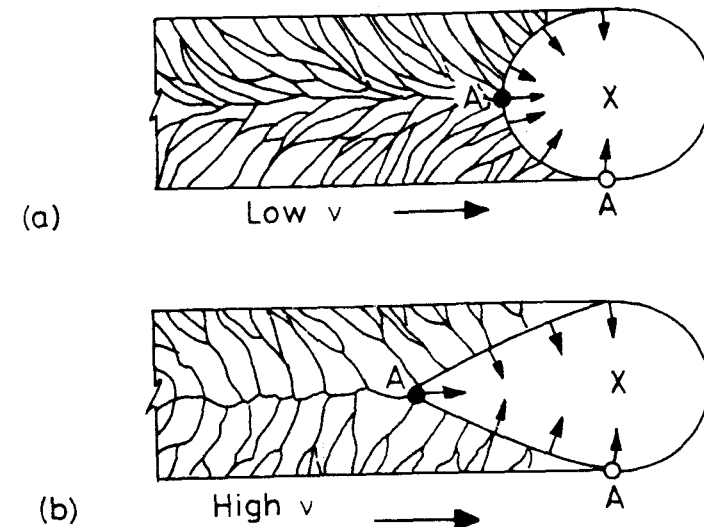


Fig. 4.49 Illustrating the effect of increasing welding speed on the shape of the melt pool and crystal growth in fusion welds.

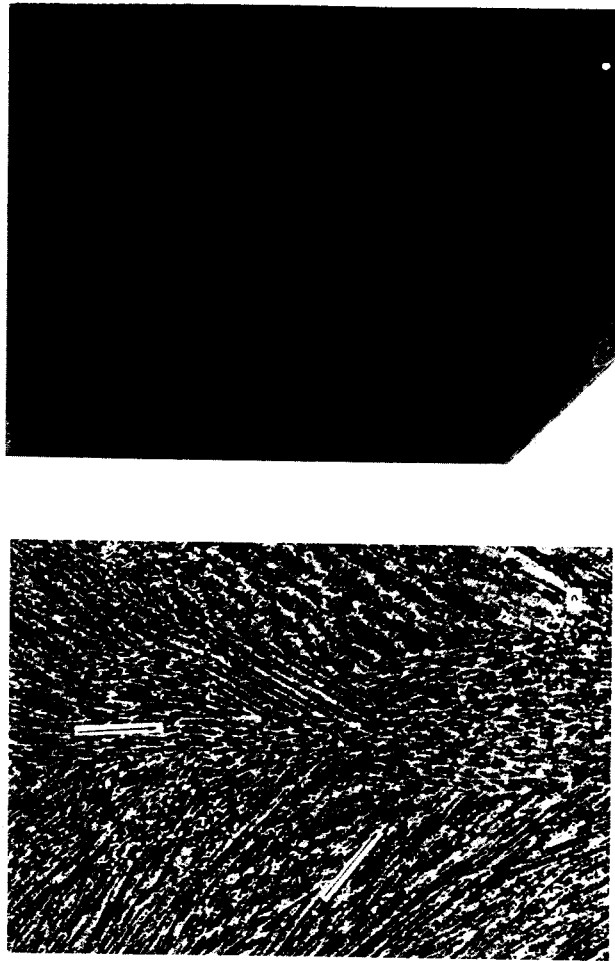


Fig. 4.50 (a) TIG weld of nickel, illustrating low crystal growth speed $\times 25$ (by Gudrun Keikkala, University of Luleå, Sweden). (b) Submerged arc weld of steel, illustrating high growth speed $\times 24$ (by H. Åström, University of Luleå, Sweden).

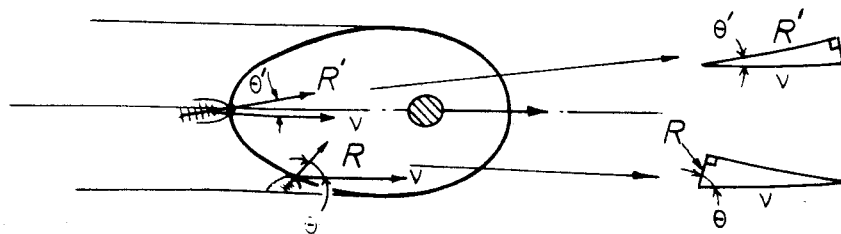


Fig. 4.51 Illustrating the relationship between crystal growth speed and welding speed in terms of rate vectors.

e.g. Fig. 4.49). Since growing crystals try to follow the steepest temperature gradients, the effect of changing the welding speed is to alter the solidification behaviour as illustrated in Fig. 4.49. As shown in 4.49b, the pear-shaped weld pool maintains fairly constant thermal gradients up to the weld centre-line, corresponding to the more angular geometry of the melt in this case. On this basis growing crystals are not required to change growth direction as at slower speeds (Fig. 4.49a). Instead, appropriately oriented crystals stabilize and widen outgrowing crystals of less favourable orientation. The crystal morphology shown in Fig. 4.49b is in fact fairly typical of the high production rate welds based on modern submerged arc welding. An example of a submerged arc weld is shown in Fig. 4.50b, and of a TIG weld of nickel in 4.50a.

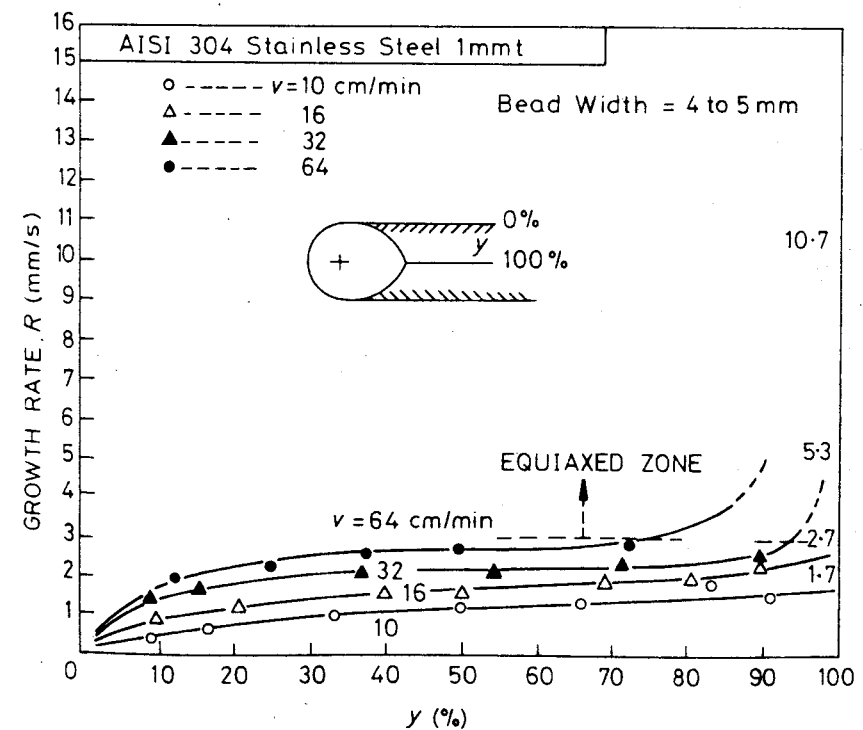
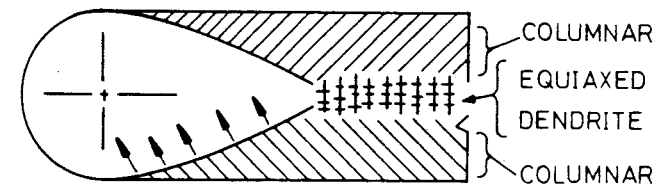


Fig. 4.52 Measurements of crystal growth rate in stainless steel as a function of per cent of weld solidified. (After T. Senda *et al.*, *Technical Report, Osaka University* 20 (1970) 932.)

While fairly linear dendritic growth is seen to predominate in this figure, it is also observed that dendrites suddenly change direction at the centre of the weld by as much as 60° . This feature of high-speed welding will now be clarified.

Geometry of Crystal Growth

Consider a welding process in which the arc is moving at a speed v . Crystal growth must occur such that it is able to keep pace with the welding speed, and this is illustrated in Fig. 4.51. It is seen that for crystal growth rate, R , to keep pace with the welding speed, v , the condition must be met that:

$$R = v \cos \theta \quad (4.49)$$

In the figure, the arrows represent vectors of speed. The vector representing the welding speed, or the speed of movement of the isotherms, is constant. On the other hand, the vector representing crystal growth rate must continuously adjust itself as growth proceeds towards the weld centre-line. It thus follows from Equation 4.49 that the solidification rate is greatest when $\theta \approx 0$, i.e. at the weld centre-line, and lowest at the weld edge where θ is a maximum. On this basis the sudden change in crystal growth direction at the weld centre line illustrated in Fig. 4.50 is associated with high growth rates as solidification attempts to keep pace with the moving arc. In addition, the initial low rates of crystal growth are associated with a relatively planar solidification front, and as the growth rate increases, the morphology of the front changes to cellular and then cellular dendritic.

An example of weld solidification rates as measured on stainless steel as a function of different welding speeds is shown in Fig. 4.52. In confirmation of Equation 4.49, it was found that completion of weld solidification ($y = 100\%$) corresponds to the highest growth rates. However, the higher welding speeds were associated with a transition from predominantly columnar crystal growth to equiaxed growth at the final stage of solidification. This transition is thought to be due to the high amounts of segregation associated with the final stages of weld solidification. This, coupled with the shallow thermal gradient at this stage leads to high degrees of constitutional supercooling and therefore the driving force for random dendritic growth to occur is large. However, it should be noted that in general dendritic and cellular substructures in welds tend to be on a finer scale than in casting, and this is mainly due to the comparatively high solidification rates of weld metal. Since higher welding speeds or thicker base metal give larger rates of solidification, it follows that the finest substructures are associated with these welds (see Equation 4.48).

When the arc is switched off at the completion of a weld run, an elliptical molten pool is left to solidify with a comparatively shallow thermal gradient. This leads to large constitutional supercooling and marked segregation. The final substructure of these weld 'craters' is thus usually equiaxed-dendritic.

Summarizing, weld solidification has the following features:

1. Solidification initially occurs epitaxially at the melted-back grains of the base metal.
2. To begin with crystal growth is relatively slow, forming first a planar and then a fine cellular substructure.
3. The intermediate stage of crystal growth is cellular-dendritic leading to coarse columnar crystal growth in the $\langle 100 \rangle$ direction in the case of cubic crystals.
4. Final solidification at the centre-line is associated with rapid crystal growth and marked segregation. Depending on welding conditions, final dendritic structure can be equiaxed.

In many ways, therefore, weld solidification and even continuous casting exhibit essentially different features to those of ingot casting (problem 4.22).

4.6 Solidification during Quenching from the Melt

The treatment of solidification presented in this chapter has been applicable for cooling rates of less than about 10^3 K/s. However, solidification can also occur at much higher rates of $10^4 - 10^7$ K/s in such processes as liquid metal atomization, melt spinning, roller-quenching or plasma spraying, as well as laser or electron beam surface treatment. By quenching melts it is possible to achieve various metastable solid states not predicted by equilibrium phase diagrams: solid phases with extended solute solubility, new metastable crystalline phases or, if the cooling rate is fast enough, amorphous metallic glasses. Crystalline solidification can occur without microsegregation or with cells or secondary dendrites spaced much more finely than in conventional solidification processes. Whether the solid is crystalline or amorphous, rapid solidification processing offers a way of producing new materials with improved magnetic or mechanical properties.

One consequence of rapid cooling can be that local equilibrium at the solid/liquid interface breaks down. Melts can solidify with no change in composition, i.e. partitionless solidification or solute trapping can occur. The thermodynamic principles involved in partitionless solidification are similar to those for the massive transformation in solids to be treated in Section 5.9.

4.7 Case Studies of some Practical Castings and Welds

4.7.1 Casting of Carbon and Low-Alloy Steels

Typical composition ranges:

C: 0.1–1.0 wt%	carbon steels	low-alloy steels	
Si: 0.1–0.4 wt%			
Mn: 0.3–1.5 wt%			
Cr: 1.0–1.6 wt%			
Ni: 1.0–3.5 wt%			
Mo: 0.1–0.1 wt%			

Casting processes: Castings, ingots, continuous casting.

Relevant phase diagrams: See Fig. 4.53.

Solidification transformations:

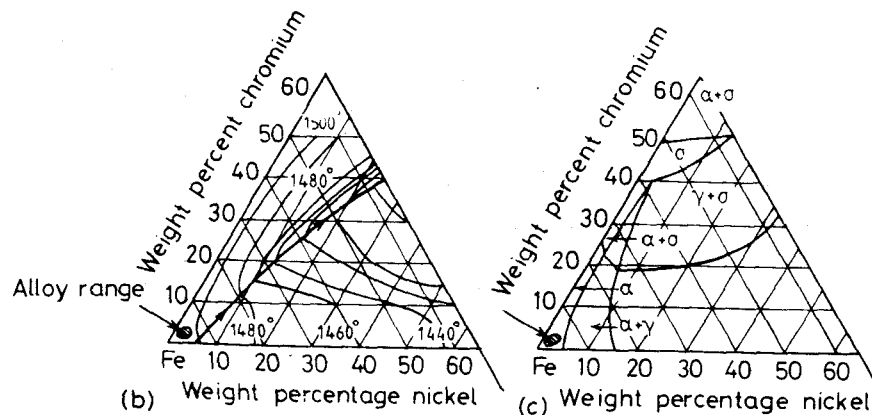
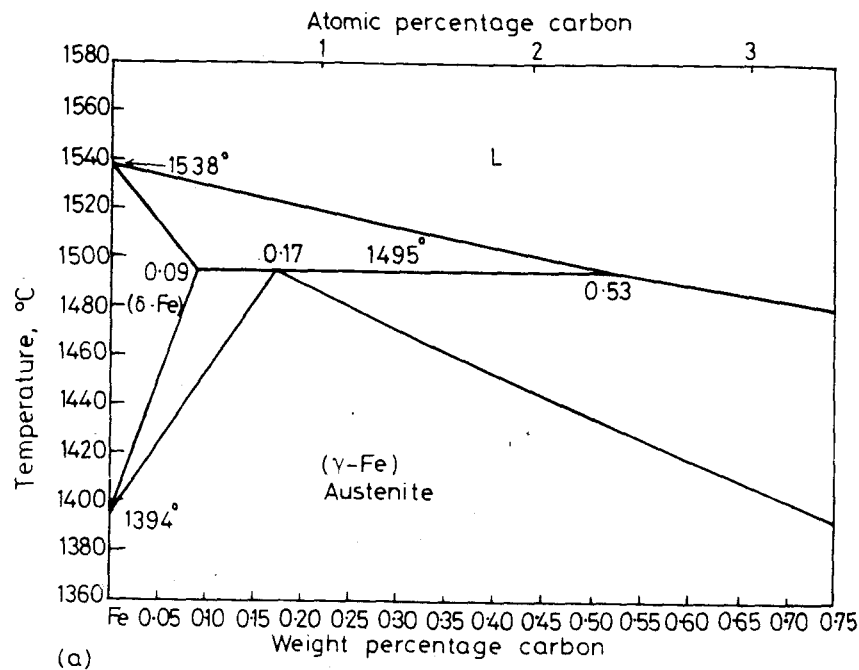
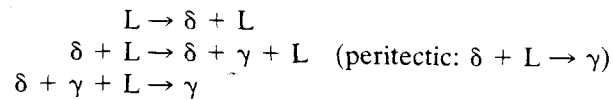
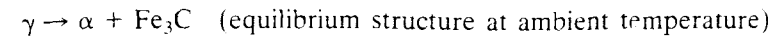


Fig. 4.53 (a) Part of the iron-carbon phase diagram. (b) Liquidus projection for the Fe-Cr-Ni system, (c) isothermal section (650 °C) for the Fe-Cr-Ni system. (From *Metals Handbook*, 8th edn., Vol. 8: American Society for Metals, 1973, pp. 276 and 425.)

Subsequent transformations:



Microstructures: See Fig. 4.54.

Comments: Figure 4.53 shows that alloying with the relatively small amounts of Ni and Cr used in low-alloy steels has little effect on solidification temperature and that the equilibrium structure of the alloy is $\alpha(+\text{Fe}_3\text{C})$. Figure 4.54a shows that quenching from the $(\gamma, \delta + L)$ field leaves a structure with considerable residual melt between solidified dendrites. As discussed earlier, in practical alloys, the presence of residual melt between dendrite arms is largely due to impurity segregation. The completely solidified structure shown in Fig. 4.54b exhibits a residual eutectic between α -Fe dendrites of $\gamma/\text{Fe}_3\text{P}/\text{Fe}_3\text{C}$, suggesting that the last liquid to solidify was rich in P and C. The retention of some γ -Fe in the eutectic is possibly due to the high carbon content of the residual iron (the solubility of P in γ -Fe is very low), which, together with the stabilizing effect of Ni, may help to retard the $\gamma \rightarrow \alpha$ transformation. Slower rates of cooling would probably reduce the amount of retained austenite still further. The presence of Mn induces the reaction: $\text{Mn} + \text{S} \rightarrow \text{MnS}$ (see Fig. 4.54b). However, this is certainly preferred to FeS , which tends to wet dendrite boundaries more extensively than MnS and is a prime cause of hot cracking.

4.7.2 Casting of High-Speed Steels

Typical composition ranges:

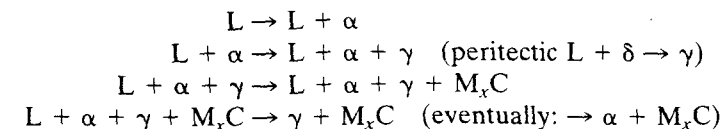
- C: 0.5–1.0 wt%
- Cr: 0.5–4.0 wt%
- Mo: 0.5–9.5 wt%
- W: 1.5–6.0 wt%
- V: 0.5–2.0 wt%

Casting processes: Ingot

Special properties: Hard, tough, wear-resistant at elevated temperatures.

Relevant phase diagrams: See Fig. 4.55.

Solidification transformations:



Microstructures: See Fig. 4.56.

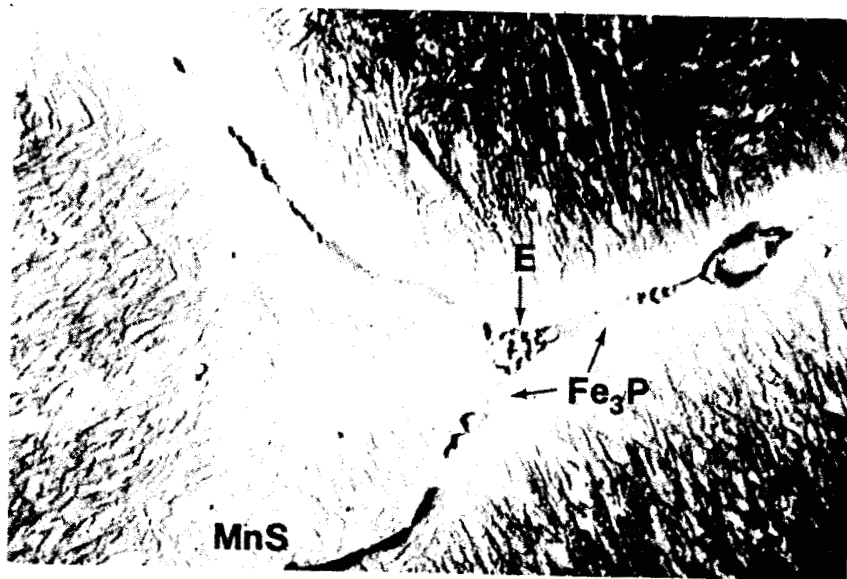


Fig. 4.54 (a) Alloy quenched from (δ , $\gamma + L$) field ($\times 25$). (b) Cooled to 20 °C. E refers to $\gamma/\text{Fe}_3\text{P}/\text{Fe}_3\text{C}$ eutectic ($\times 1000$). (From *Guide to the Solidification of Steels*, Jernkontoret, Stockholm, 1977.)

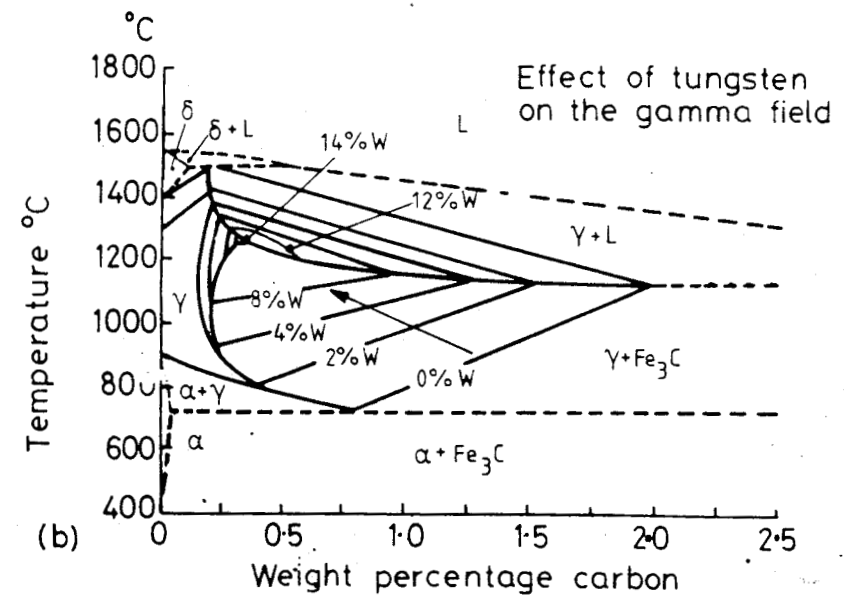
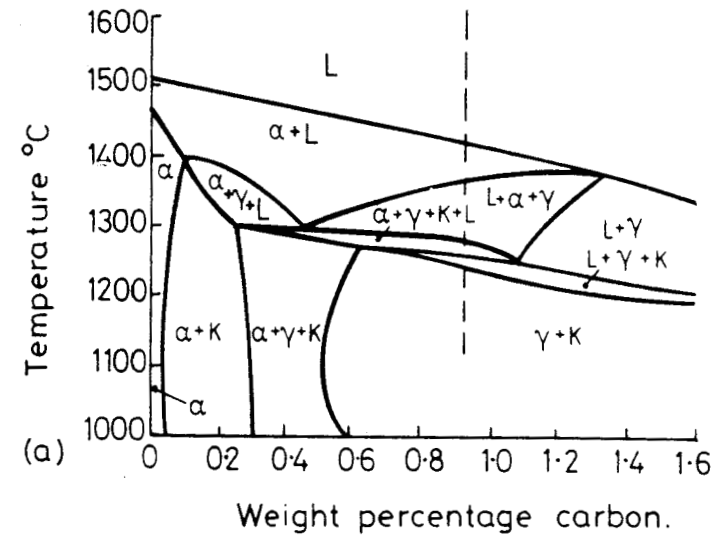


Fig. 4.55 (a) Phase diagram for steel with approx 4wt% Cr, 5wt% Mo, 6wt% W and 2wt% V (after E. Horn and H. Brandis, *DEW-Techn. Ber.* 11 (1971) 147). (b) Effect of W on γ field of steel (from *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 416).

Comments: Reference to the phase diagrams (Fig. 4.55) shows that the presence of W, V and the other main alloying elements produces a cascade of polyphase fields during cooling of these castings. Solidification occurs initially with the formation of α dendrites, but the γ fields are so extensive that rapid quenching from the $\alpha + L$ field can not suppress the nucleation and growth of austenite (Fig. 4.56a). As expected, the C segregates strongly when α forms, but W, Cr and V are not expected to segregate so markedly in α -Fe. It seems likely that the early formation of γ through the reaction: $\alpha + L \rightarrow \gamma + L$ in these castings is the main cause of the extensive segregation of W, Cr and V as observed in Fig. 4.56b and c. As seen in Fig. 4.56b, it is possible that the $\alpha \rightarrow \gamma$ reaction occurs through the rejection of dissolved M back to the melt. Reference to the Fe-Cr, Fe-V, and Fe-W binary phase diagrams shows in all cases very low high-temperature solubility of these elements in γ -Fe. The resulting as-solidified structure (Fig. 4.56c) thus consists of α dendrites (following the $\gamma \rightarrow \alpha$ solid-state transformation during cooling) with marked

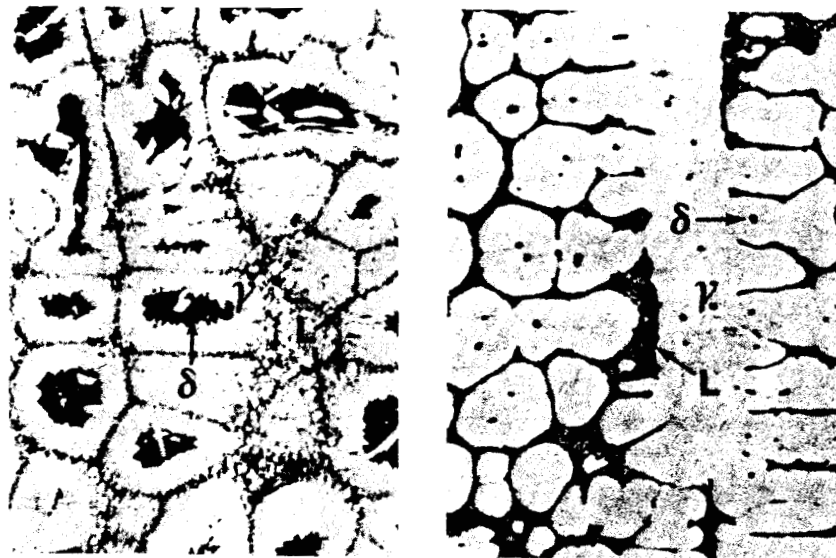


Fig. 4.56 (a) Quenched from the $(L + \alpha \text{ (or } \delta) + \gamma)$ field at 1335°C ($\times 150$). (b) Quenched from 1245°C ($\times 150$). (c) Same alloy after mechanical and thermal heat treatments ($\times 750$). (a-c from *A Guide to the Solidification of Steel*, Jernkontoret, Stockholm, 1977. and d from *Metals Handbook*, 8th edn., Vol. 7, American Society for Metals, 1972, p. 121.)

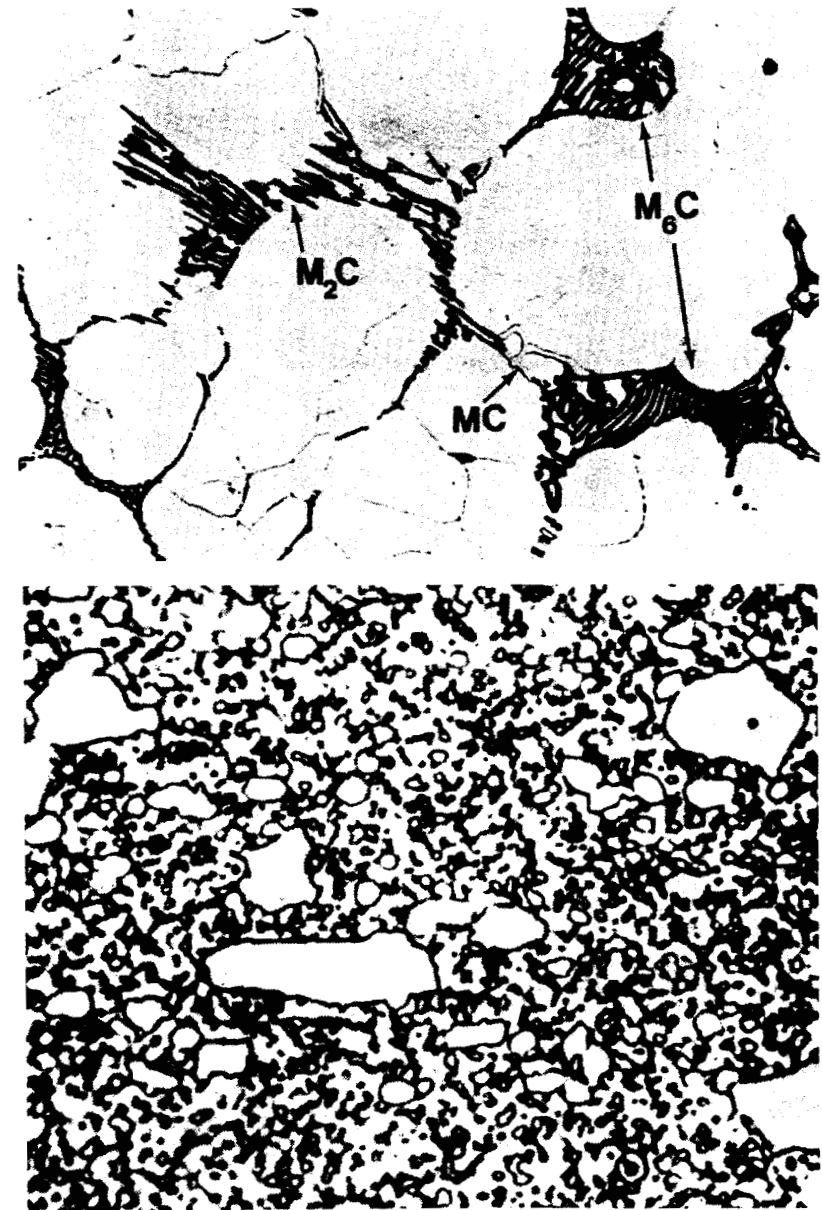


Fig. 4.56b & c (continued)

interdendritic segregation. The latter appears in the form of a γ/M_xC eutectic, where M_xC refers to mixed carbides of WC, Cr_2C , VC, etc. The final structure of this type of tool steel (Fig. 4.56d) is only reached after further extensive plastic working to break up the eutectic, followed by austenitizing and double-tempering treatments.

4.7.3 Stainless Steel Weld Metal

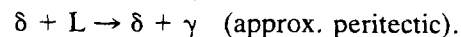
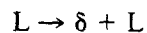
Electrode composition range:

Cr:	17–19 wt%
Ni:	8–10 wt%
C:	0.05–0.1 wt%
Si:	0.5–1.0 wt%
Mn:	0.5–1.5 wt%
P:	traces
S:	

Welding process: Manual metal arc, gas metal arc.

Relevant phase diagrams: See Fig. 4.57.

Phase transformations:



Microstructures: See Fig. 4.58.

Comments:

(a) *Phase equilibria.* Figure 4.57a is the 18% Cr vertical section of the Fe–Cr–Ni ternary diagram. The effect of these and other alloying elements on the final microstructure, assuming fairly high quench rates typical of welding, can be predicted with the help of the Schaeffler diagram 4.57b. From this diagram it can be seen that if the Ni or Cr contents are reduced much below the nominal analyses given, there is a risk that martensite forms. The most important feature of these alloys with respect to welding, is that some δ -Fe is retained even at ambient temperatures (see 4.57b).

There has been much discussion in the literature as to which phase solidifies first after welding. According to Fig. 4.57a it appears that solidification should initiate with the nucleation of δ -Fe. However, if the base metal is fully austenitic at the transition zone, this phase should nucleate first because of the requirement of epitaxial growth (see previous discussion). Unfortunately the situation is complicated in practice by the presence of carbon and nitrogen, both of which tend to move the peritectic composition towards higher Ni content. An example of the effect of N being admitted to the weld pool is shown in Fig. 4.58a illustrating a single run weld which has remained fully austenitic. The result of this is fairly catastrophic, causing hot cracking at the austenite grain boundaries due to increased sulphur and phosphorus segregation in the austenite during solidification.

(b) *Microstructure.* It is thought that the first phase to solidify in this alloy is δ -Fe, enriched in Cr and impoverished in Ni, the tendency in either case being to stabilize the ferrite. Further cooling causes γ -Fe to nucleate in the Ni-rich liquid between the δ -Fe dendrites. With the development of this duplex $\gamma + \delta$ structure the peritectic reaction: $L \rightarrow \gamma + \delta$ continues to completion. Cooling of the weld metal to ambient temperature causes the γ -Fe phase to grow

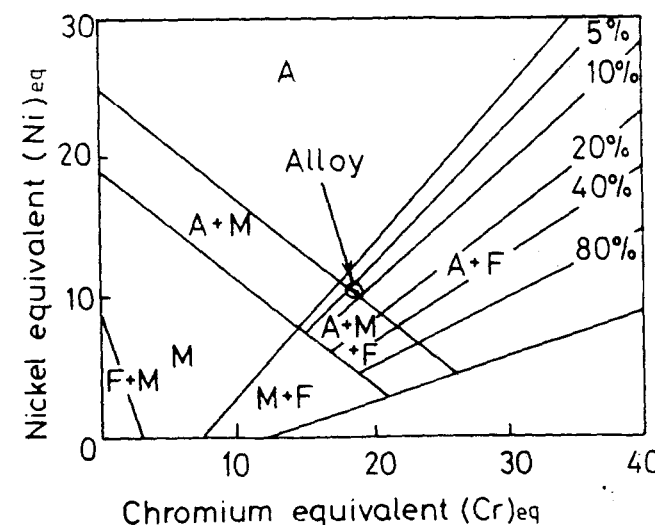
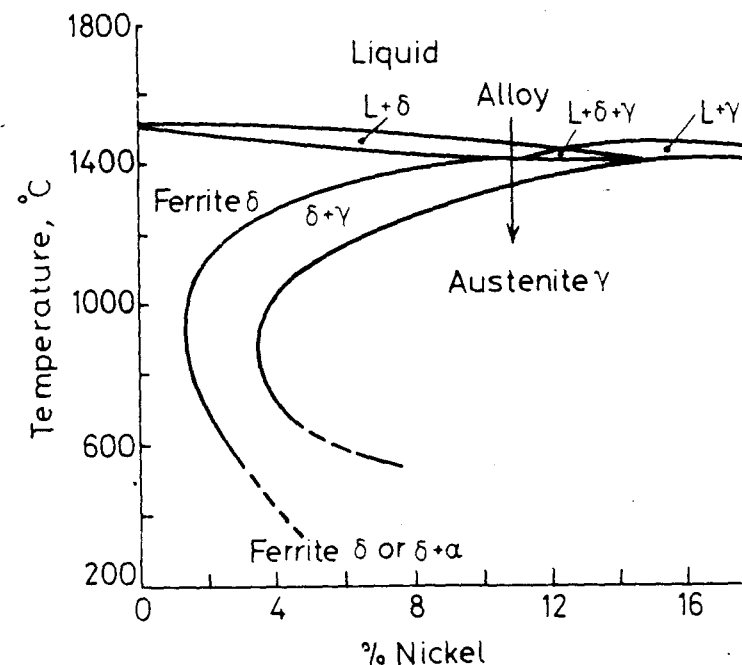


Fig. 4.57 (a) 18% Cr section of the Fe–Cr–Ni system. (b) Schaeffler diagram indicating the alloy concerned. $(Ni)_{eq} = \%Ni + 30 \times \%C + 0.5 \times \%Mn$. $(Cr)_{eq} = \%Cr + \%Mo + 1.5 \times \%Si + 0.5 \times \%Nb$. A, austenite; F, ferrite; M, martensite. (After R.J. Castro and J.J. de Cadenet, *Welding Metallurgy of Stainless and Heat Resistant Steels*, Cambridge University Press, Cambridge, 1974.)

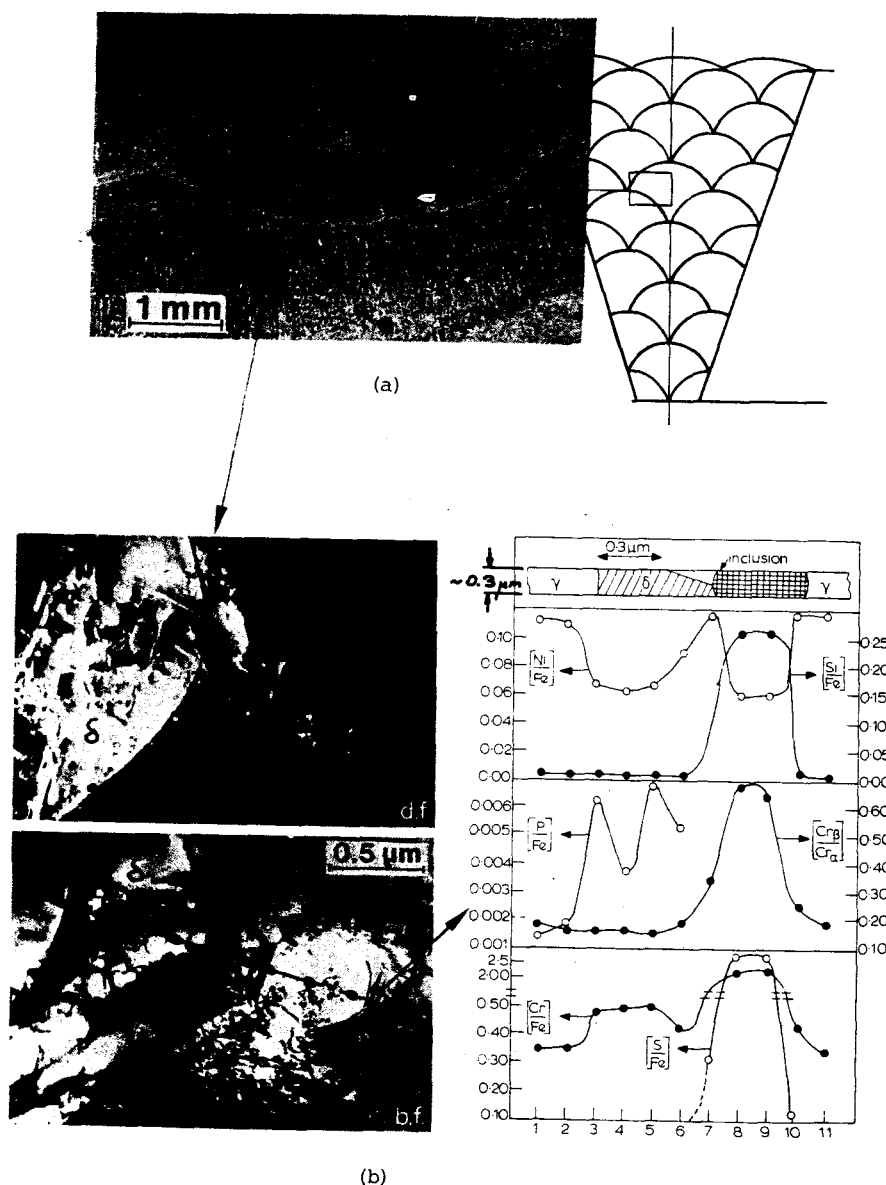


Fig. 4.58 (a) Illustrating hot cracking in the austenite region of a duplex stainless steel weld deposit. (b) STEM-EDX microanalysis of γ and δ -Fe and an inclusion. (After H. Åström *et al.*, *Metal Science Journal*, July 1976, p. 225.)

at the expense of δ -Fe until only a fine network of δ -Fe remains. The STEM-based X-ray spectrometer microanalysis of the γ , δ and inclusion phases (Fig. 4.58b) indicates that the Cr-rich ferrite has dissolved the phosphorus (one of the danger elements in hot cracking), while the sulphur is bound up in the inclusion. In this respect Mn has a double role: both as a

deoxidizer and to absorb S through forming MnS. If the weld solidifies directly to γ -Fe, all the Mn remains in solution and thus cannot prevent FeS forming at cell boundaries. The fine duplex $\gamma + \delta$ structure of stainless steel welds thus refines and strengthens the microstructure, and effectively renders S and P harmless. It should be pointed out, however, that the 6–8 vol% retained δ -Fe at ambient temperature should not be exceeded, since higher volume fractions reduce the ductility and toughness of this alloy. In this respect, the Shaeffler diagram (Fig. 4.57b) is a useful guide for estimating δ -Fe as a function of equivalent Cr and Ni content. The amount of δ -Fe can also be measured magnetically or metallographically. If the presence of nitrogen is to be accounted for, a modified form of Shaeffler diagram (the DeLong diagram) can be employed (see, e.g. Castro and Cadenet, 1974)¹⁸.

References

1. D. Turnbull and R.E. Cech, *Journal of Applied Physics*, **21**:804 (1950).
2. R.W. Cahn, *Nature*, **273**:491 (1978).
3. J. Frenkel, *Kinetic Theory of Liquids*, Dover, New York, 1955.
4. R.M.J. Cotterill *et al.*, *Phil. Mag.*, **31**:245 (1975).
5. Reasons for this crystallographic feature of dendrite growth are discussed for example in B. Chalmers, *Principles of solidification*, Wiley, 1964, p. 116.
6. V.G. Smith *et al.*, *Canadian Journal of Physics*, **33**:723 (1955).
7. W.G. Pfann, *Zone Melting*, 2nd edition, John Wiley & Sons, New York, 1966.
8. K.A. Jackson and J.D. Hunt, *Acta Metallurgica*, **13**:1212 (1965).
9. For a review of eutectic solidification see R. Elliot, 'Eutectic solidification', Review 219 in *International Metals Reviews*, September 1977, p. 161.
10. See M.C. Flemings, *Solidification Processing*, McGraw-Hill, 1974, p. 101.
11. For a complete treatment of these effects see K.A. Jackson and J.D. Hunt, *Transactions of the Metallurgical Society of AIME*, **236**:1129 (1966).
12. M.H. Burden and J.D. Hunt, 'The extent of the eutectic range', *Journal of Crystal Growth*, **22**:328 (1974).
13. F.D. Lemkey and N.J. Salikind in H.S. Peiser (Ed.), *Crystal Growth*, Pergamon, Oxford, 1967, p. 171.
14. M.C. Flemings, *Solidification Processing*, McGraw-Hill, 1974, p. 177.
15. A. Ohno, *The Solidification of Metals*, Chijin Shokan, Tokyo, 1976, p. 69.
16. For a detailed treatment see M.C. Flemings, 'Principles of control of soundness and homogeneity of large ingots', *Scandinavian Journal of Metallurgy*, **5**:1 (1976).

17. R.J. Jesseman, *Microalloying '75*, (J. Crane, ed.) Union Carbide Corporation, Washington DC, 1975, Vol 3, p. 74.
18. R.J. Castro and J.J. de Cadenet, *Welding Metallurgy of Stainless and Heat Resistant Steels*, Cambridge University Press, 1974.
19. D. Rosenthal, *Transactions of the American Society of Mechanical Engineers*, **68**:849 (1946).

Further Reading

- R.W. Cahn and P. Haasen (Eds.), 'Physical metallurgy', Chapter 9, H. Biloni, *Solidification*, and Chapter 28, R.W. Cahn, *Alloys Quenched from the Melt*. North-Holland, 1983.
- G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972.
- B. Chalmers, *Principles of Solidification*, John Wiley and Sons, 1964.
- G.J. Davies and P. Garland, 'Solidification structures and properties of fusion welds', *Int. Met. Rev.*, June, 1975.
- S.A. David and J.M. Vitek, 'Correlation between solidification structures and weld microstructures', *International Materials Review*, **34**: 213-45 (1989).
- K.E. Easterling, *Introduction to the Physical Metallurgy of Welding*, Butterworth-Heinemann, London, 2nd edn, 1992.
- M.C. Flemings, *Solidification Processing*, McGraw-Hill, 1974.
- W. Kurz and D.J. Fisher, *Fundamentals of Solidification*, Trans. Tech. Publications, Clausthal-Zellerfeld, Germany, 1984.
- H. Ohno, *The Solidification of Metals*, Chijin Shokan, 1976.
- Solidification and Casting of Metals*, book 192, The Metals Society, London, 1979.

Exercises

- 4.1 Show that differentiation of Equation 4.4 leads to Equations 4.5 and 4.6.
- 4.2 Use Equations 4.4 and 4.10 to estimate the number of crystal-like clusters in 1 mm^3 of copper at its melting point for spherical clusters containing (a) 10 atoms, (b) 60 atoms. What volume of liquid copper is likely to contain one cluster of 100 atoms? The atomic volume of liquid copper is $1.6 \times 10^{-29} \text{ m}^3$, $\gamma_{\text{SL}} = 0.177 \text{ J m}^{-2}$, $k = 1.38 \times 10^{-23} \text{ J K}^{-1}$, $T_{\text{m}} = 1356 \text{ K}$.
- 4.3 Why does r_{max} in Fig. 4.5 vary with ΔT ?
- 4.4 Calculate the homogeneous nucleation rate in liquid copper at undercoolings of 180, 200 and 220 K, using the following data:

$$L = 1.88 \times 10^9 \text{ J m}^{-3}, \quad T_{\text{m}} = 1356 \text{ K}, \quad \gamma_{\text{SL}} = 0.177 \text{ J m}^{-2}, \\ f_0 = 10^{11} \text{ s}^{-1}, \quad C_0 = 6 \times 10^{28} \text{ atoms m}^{-3}, \quad k = 1.38 \times 10^{-23} \text{ J K}^{-1}.$$

- 4.5 Show that Equation 4.23 applies to homogeneous nucleation and heterogeneous nucleation on a flat mould wall.
- 4.6 Show that Equation 4.16 follows from 4.15 using the following relationships for a spherical cap:

$$A_{\text{SL}} = 2\pi r^2 (1 - \cos \theta)$$

$$A_{\text{SM}} = \pi r^2 \sin^2 \theta$$

$$V_{\text{S}} = \pi r^3 (2 + \cos \theta)(1 - \cos \theta)^2/3$$

- 4.7 Of what importance is the angle of a mould-wall crack in heterogeneous nucleation? Of what importance is the width of the crack at its mouth?
- 4.8 Under what conditions can solid metal be retained in a mould-wall crevice above T_{m} ?
- 4.9 If a single crystal is melted by heating to slightly above its melting point and then cooled it subsequently solidifies with the previous orientation. Likewise a polycrystalline specimen reverts to its original grain size. Can you suggest an explanation for this effect? (See B. Chalmers, *Principles of Solidification*, Wiley, 1964, p. 85).
- 4.10 (a) Show that surface melting is to be expected below T_{m} in gold (1336 K) given $\gamma_{\text{SL}} = 0.132$, $\gamma_{\text{LV}} = 1.128$, $\gamma_{\text{SV}} = 1.400 \text{ J m}^{-2}$.
(b) Given that the latent heat of fusion of gold is $1.2 \times 10^9 \text{ J m}^{-3}$ estimate whether sensible liquid layer thicknesses are feasible at measurably lower temperatures than T_{m} .
- 4.11 Use nucleation theory to derive quantitative expressions for the velocity of an atomically smooth interface as a function of undercooling (a) for repeated surface nucleation, (b) for spiral growth. (See Burton *et al.*, *Philosophical Transactions*, **A243**:299 (1950)).
- 4.12 Draw diagrams to show how the solid/liquid interface temperature varies as a function of position along the bar for Figs. 4.20, 4.21 and 4.22.
- 4.13 Draw figures corresponding to Figs. 4.21 and 4.22 for a dilute binary alloy with $k > 1$.
- 4.14 Show that Equation 4.35 satisfies 4.34.
- 4.15 The Al-Cu phase diagram is similar to that shown in Fig. 4.19 with $T_{\text{m}}(\text{Al}) = 660^\circ\text{C}$, $T_{\text{E}} = 548^\circ\text{C}$, $X_{\text{max}} = 5.65 \text{ wt\%}$, and $X_{\text{E}} = 33 \text{ wt\% Cu}$. The diffusion coefficient for the liquid $D_{\text{L}} = 3 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$. If an Al-0.5 wt% Cu alloy is solidified with no convection and a planar solid/liquid interface at $5 \mu\text{m s}^{-1}$:
(a) What is the interface temperature in the steady state?
(b) What is the thickness of the diffusion layer?
(c) What temperature gradient will be required to maintain a planar interface?

- (d) Answer (a), (b) and (c) for an Al-2 wt% Cu alloy solidified under the same conditions.
- 4.16 (a) Using Equation 4.33 and the data in problem 4.15 plot the variation of copper concentration along a unidirectionally solidified bar of an Al-2 wt% Cu alloy assuming no diffusion in the solid and perfect mixing in the liquid.
- (b) What fraction of the bar will solidify to a eutectic structure?
- (c) How much eutectic would form in an Al-0.5 wt% Cu alloy solidified under the same conditions?
- 4.17 Explain the experimental observation that in the presence of a convection current cells grow upstream.
- 4.18 Sketch a possible solidification-front structure for the solidification of an Fe-0.25 wt% C alloy in a shallow temperature gradient. Consider the temperature range 1440–1540 °C. Assume very rapid diffusion of carbon in δ -Fe.
- 4.19 Show that the condition $\lambda = 2\lambda^*$ gives (i) the maximum eutectic growth rate for a given undercooling, and (ii) a minimum undercooling for a given growth rate (Equation 4.43).
- 4.20 Calculate the depression of the eutectic temperature for a lamellar eutectic with $\lambda = 0.2 \mu\text{m}$ and $\lambda = 1.0 \mu\text{m}$, if $\gamma_{\alpha\beta} = 400 \text{ mJ m}^{-2}$, $\Delta H/V_m = 800 \times 10^6 \text{ J m}^{-3}$, $T_E = 1000 \text{ K}$.
- 4.21 If it is assumed that the choice of a rod or lamellar eutectic is governed by the minimization of the total α/β interfacial energy it can be shown that for a given λ there is a critical volume fraction of the β phase (f_c) below which β should be rod like, and above which it should be lamellar. Assuming the rods are hexagonally arranged and that $\gamma_{\alpha\beta}$ is isotropic, calculate the value of f_c .
- 4.22 Compare the processes of ingot casting and weld solidification, and show they are in many ways quite different solidification processes. How would you compare continuous casting in this respect?
- 4.23 What is the influence of welding speed on the solidification structure of welds? How is welding speed likely to affect segregation problems?

5

Diffusional Transformations in Solids

The majority of phase transformations that occur in the solid state take place by thermally activated atomic movements. The transformations that will be dealt with in this chapter are those that are induced by a change of temperature of an alloy that has a fixed bulk composition. Usually we will be concerned with the transformations caused by a temperature change from a single-phase region of a (binary) phase diagram to a region where one or more other phases are stable. The different types of phase transformations, that are possible can be roughly divided into the following groups: (a) precipitation reactions, (b) eutectoid transformations, (c) ordering reactions, (d) massive transformations, and (e) polymorphic changes. Figure 5.1 shows several different types of binary phase diagrams that are representative of these transformations.

Precipitation transformations can be expressed in reaction terms as follows



where α' is a metastable *supersaturated solid solution*, β is a stable or metastable precipitate, and α is a more stable solid solution with the same crystal structure as α' , but with a composition closer to equilibrium, see Fig. 5.1a.

Eutectoid transformations involve the replacement of a metastable phase (γ) by a more stable mixture of two other phases ($\alpha + \beta$) and can be expressed as



This reaction is characteristic of phase diagrams such as that shown in Fig. 5.1b.

Both precipitation and eutectoid transformations involve the formation of phases with a different composition to the matrix and therefore long-range diffusion is required. The remaining reaction types can, however, proceed without any composition change or long-range diffusion. Figure 5.1c shows phase diagrams where *ordering reactions* can occur. In this case the reaction can be simply written



In a *massive* transformation the original phase decomposes into one or more new phases which have the same composition as the parent phase, but different crystal structures. Figure 5.1d illustrates two simple examples of the type



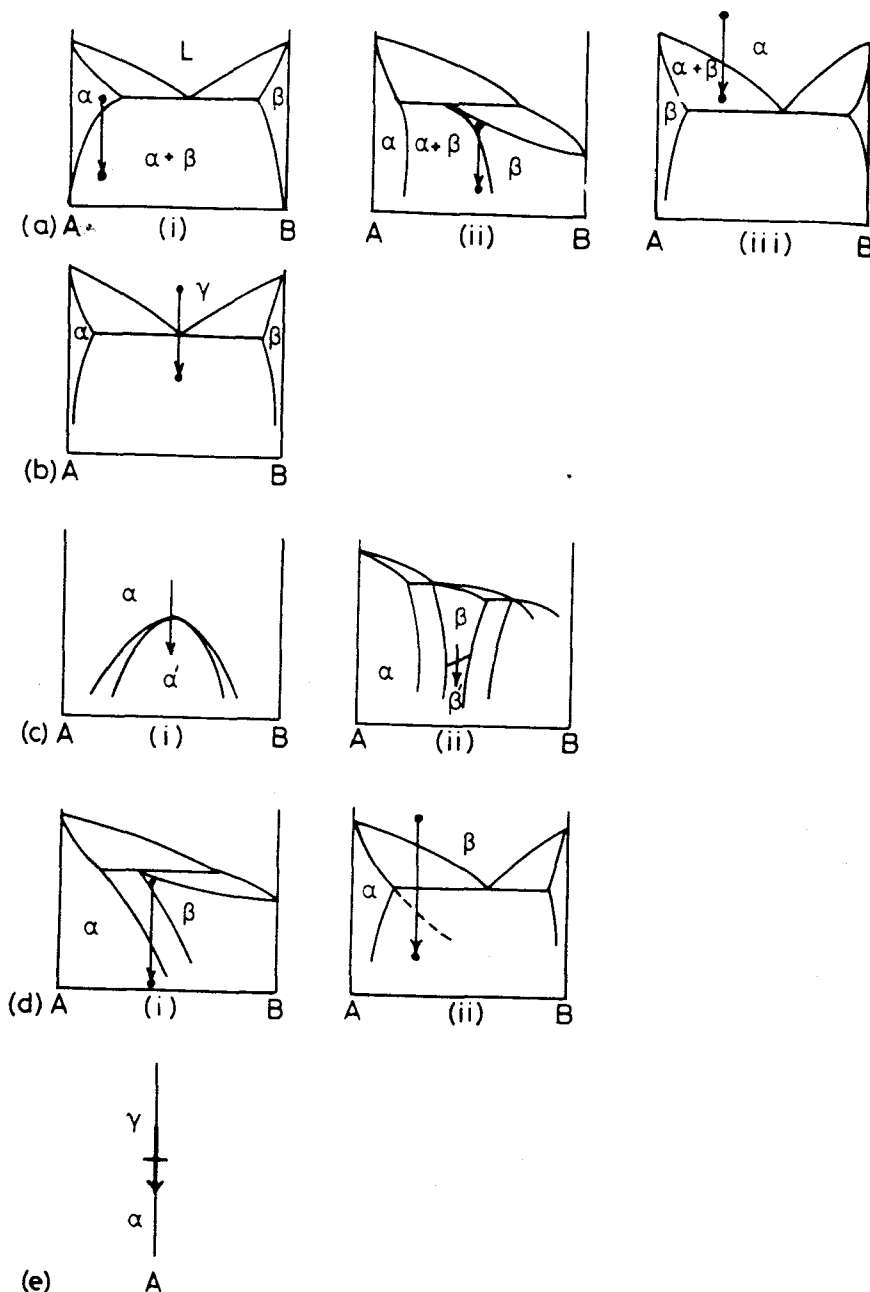


Fig. 5.1 Examples of different categories of diffusional phase transformations: (a) precipitation; (b) eutectoid; (c) ordering; (d) massive; (e) polymorphic (single component).

where only one new phase results. Note that the new β phase can either be stable (Fig. 5.1d(i)) or metastable (Fig. 5.1d(ii)).

Polymorphic transformations occur in single component systems when different crystal structures are stable over different temperature ranges, Fig. 5.1e. The most well-known of these in metallurgy are the transformations between fcc- and bcc-Fe. In practice, however, such transformations are of little practical interest and have not been extensively studied.

Apart from a few exceptions the above transformations all take place by diffusional nucleation and growth. As with solidification, nucleation is usually heterogeneous, but for the sake of simplicity let us begin by considering homogeneous nucleation.

5.1 Homogeneous Nucleation in Solids

To take a specific example consider the precipitation of B-rich β from a supersaturated A-rich α solid solution as shown in Fig. 5.1a(i). For the nucleation of β , B-atoms within the α matrix must first diffuse together to form a small volume with the β composition, and then, if necessary, the atoms must rearrange into the β crystal structure. As with the liquid $\rightarrow$ solid transformation an α/β interface must be created during the process and this leads to an activation energy barrier.

The free energy change associated with the nucleation process will have the following three contributions.

1. At temperatures where the β phase is stable, the creation of a volume V of β will cause a volume free energy *reduction* of $V\Delta G_v$.
2. Assuming for the moment that the α/β interfacial energy is isotropic the creation of an area A of interface will give a free energy *increase* of $A\gamma$.
3. In general the transformed volume will not fit perfectly into the space originally occupied by the matrix and this gives rise to a misfit strain energy ΔG_s per unit volume of β . (It was shown in Chapter 3 that, for both coherent and incoherent inclusions, ΔG_s is proportional to the volume of the inclusion.) Summing all of these gives the total free energy change as

$$\Delta G = -V\Delta G_v + A\gamma + V\Delta G_s \quad (5.5)$$

Apart from the misfit strain energy term, Equation 5.5 is very similar to that derived for the formation of a solid nucleus in a liquid. With solid/liquid interfaces γ can be treated as roughly the same for all interfaces, but for nucleation in solids γ can vary widely from very low values for coherent interfaces to high values for incoherent interfaces. Therefore the $A\gamma$ term in Equation 5.5 should really be replaced by a summation over all surfaces of the nucleus $\sum \gamma_i A_i$.

If we ignore the variation of γ with interface orientation and assume the nucleus is spherical with a radius of curvature r Equation 5.5 becomes

$$\Delta G = -\frac{4}{3}\pi r^3(\Delta G_v - \Delta G_s) + 4\pi r^2\gamma \quad (5.6)$$

This is shown as a function of r in Fig. 5.2. Note that the effect of the misfit strain energy is to reduce the effective driving force for the transformation to $(\Delta G_v - \Delta G_s)$. Similar curves would in fact be obtained for any nucleus shape as a function of its size. Differentiation of Equation 5.6 yields

$$r^* = \frac{2\gamma}{(\Delta G_v - \Delta G_s)} \quad (5.7)$$

$$\Delta G^* = \frac{16\pi\gamma^3}{3(\Delta G_v - \Delta G_s)^2} \quad (5.8)$$

which is very similar to the expressions for solidification, except now the chemical driving force ΔG_v is reduced by a positive strain energy term.

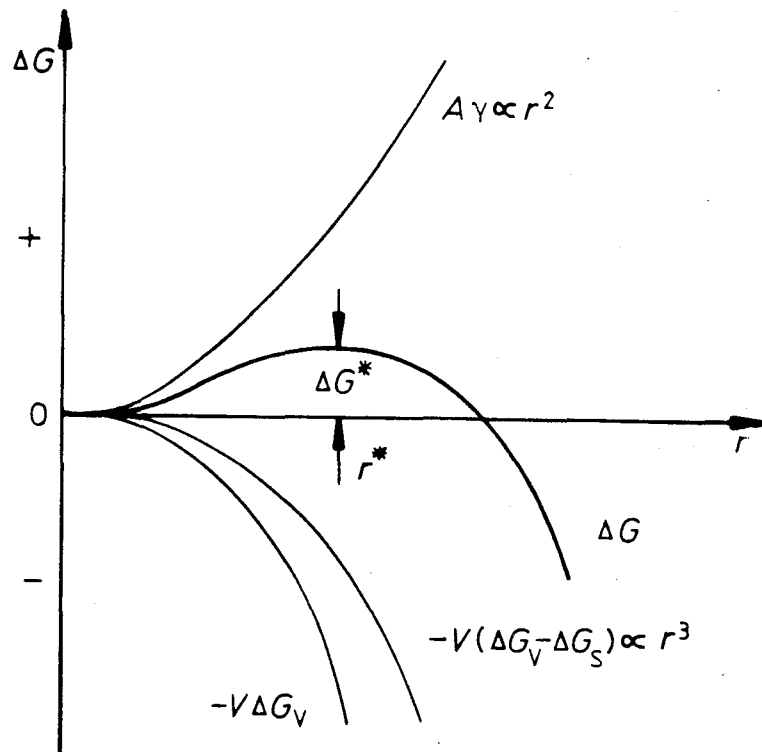


Fig. 5.2 The variation of ΔG with r for a homogeneous nucleus. There is an activation energy barrier ΔG^* .

As discussed in Chapter 4 the concentration of critical-sized nuclei C^* will be given by

$$C^* = C_0 \exp(-\Delta G^*/kT) \quad (5.9)$$

where C_0 is the number of atoms per unit volume in the phase. If each nucleus can be made supercritical at a rate of f per second the homogeneous nucleation rate will be given by

$$N_{\text{hom}} = fC^* \quad (5.10)$$

f depends on how frequently a critical nucleus can receive an atom from the α matrix. This will depend on the surface area of the nucleus and the rate at which diffusion can occur. If the activation energy for atomic migration is ΔG_m per atom, f can be written as $\omega \exp(-\Delta G_m/kT)$ where ω is a factor that includes the vibration frequency of the atoms and the area of the critical nucleus. The nucleation rate will therefore be of the form

$$N_{\text{hom}} = \omega C_0 \exp\left(-\frac{\Delta G_m}{kT}\right) \exp\left(-\frac{\Delta G^*}{kT}\right) \quad (5.11)$$

This is essentially identical to Equation 4.12 except that the temperature dependence of f has been taken into account. In order to evaluate this equation as a function of temperature ω and ΔG_m can be assumed to be constant, but ΔG^* will be strongly temperature dependent. The main factor controlling ΔG^* is the driving force for precipitation ΔG_v . Equation 5.8. Since composition is variable the magnitude of ΔG_v must be obtained from the free energy–composition diagram.

If the alloy X_0 in Fig. 5.3, is solution treated at T_1 and then cooled rapidly to T_2 it will become supersaturated with B and will try to precipitate β . When the transformation to $\alpha + \beta$ is complete the free energy of the alloy will have decreased by an amount ΔG_0 per mole as shown in Fig. 5.3b. ΔG_0 is therefore the total driving force for the transformation. However, it is not the driving force for *nucleation*. This is because the first nuclei to appear do not significantly change the α composition from X_0 . The free energy released per mole of nuclei formed can be obtained as follows.

If a small amount of material with the nucleus composition (X_B^β) is removed from the α phase, the total free energy of the system will *decrease* by ΔG_1 where

$$\Delta G_1 = \mu_A^\alpha X_A^\beta + \mu_B^\alpha X_B^\beta \quad (\text{per mol } \beta \text{ removed}) \quad (5.12)$$

This follows simply from the definition of chemical potential given by Equation 1.29. ΔG_1 is a quantity represented by point P in Fig. 5.3b. If these atoms are now rearranged into the β crystal structure and replaced, the total free energy of the system will *increase* by an amount

$$\Delta G_2 = \mu_A^\beta X_A^\beta + \mu_B^\beta X_B^\beta \quad (\text{per mol } \beta \text{ formed}) \quad (5.13)$$

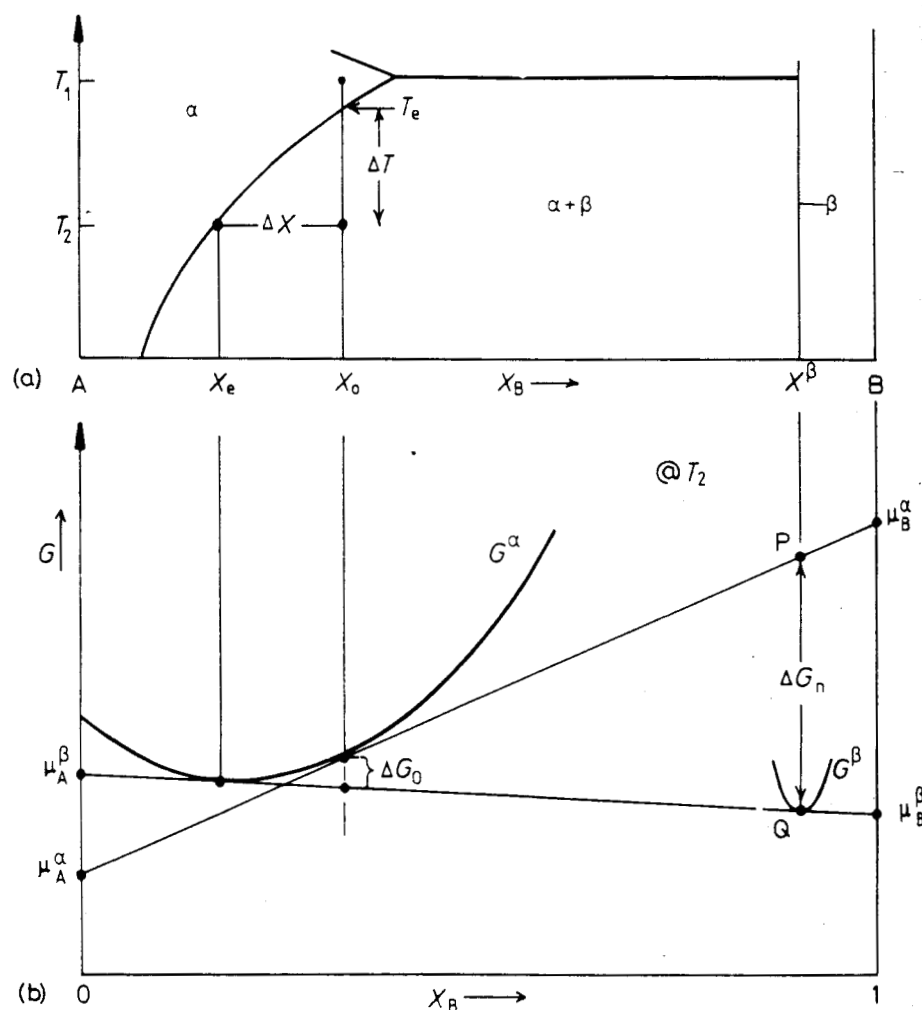


Fig. 5.3 Free energy changes during precipitation. The driving force for the first precipitates to nucleate is $\Delta G_n = \Delta G_v V_m$. ΔG_0 is the total decrease in free energy when precipitation is complete and equilibrium has been reached.

which is given by point Q. Therefore the driving force for nucleation

$$\Delta G_n = \Delta G_2 - \Delta G_1 \quad \text{per mol of } \beta \quad (5.14)$$

which is just the length PQ in Fig. 5.3b. The *volume* free energy decrease associated with the nucleation event is therefore simply given by

$$\Delta G_v = \frac{\Delta G_n}{V_m} \quad \text{per unit volume of } \beta \quad (5.15)$$

where V_m is the molar volume of β . For dilute solutions it can be shown that

approximately

$$\Delta G_v \propto \Delta X \quad (5.16)$$

where

$$\Delta X = X_0 - X_e \quad (5.17)$$

From Fig. 5.3a therefore it can be seen that the driving force for precipitation increases with increasing undercooling (ΔT) below the equilibrium solvus temperature T_e .

It is now possible to evaluate Equation 5.11 for alloy X_0 as a function of temperature. The variation of ΔG_v with temperature is shown schematically in Fig. 5.4b. After taking into account the misfit strain energy term ΔG_s the

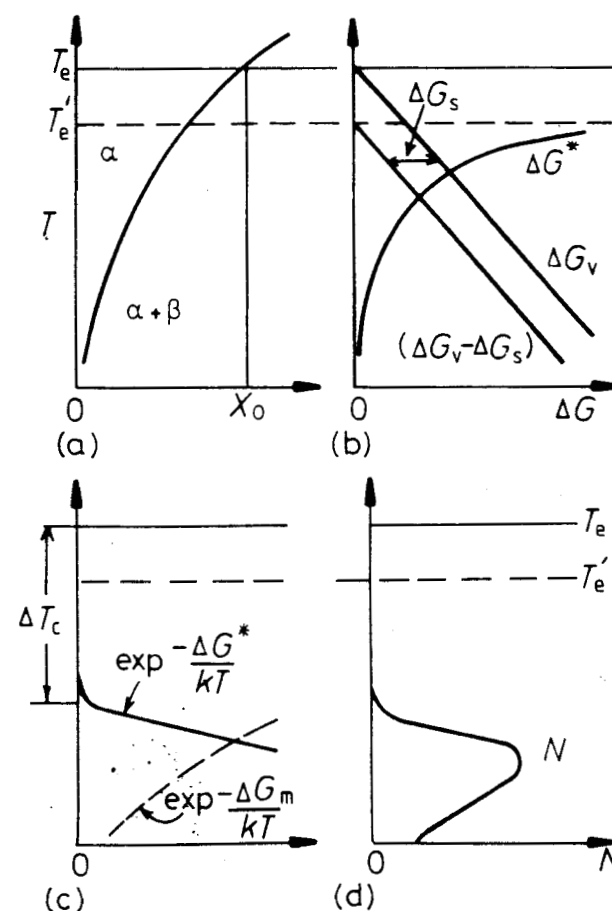


Fig. 5.4 How the rate of homogeneous nucleation varies with undercooling for alloy X_0 . (a) The phase diagram. (b) The effective driving force ($\Delta G_v - \Delta G_s$) and the resultant energy barrier ΔG^* . (c) The two exponential terms that determine N as shown in (d).

effective driving force becomes $(\Delta G_v - \Delta G_s)$ and the effective equilibrium temperature is reduced to T'_e . Knowing $(\Delta G_v - \Delta G_s)$ the activation energy ΔG^* can be calculated from Equation 5.8 as shown. Figure 5.4c shows the two exponential terms in Equation 5.11; $\exp(-\Delta G^*/kT)$ is essentially the potential concentration of nuclei and, as with nucleation in liquids, this is essentially zero until a critical undercooling ΔT_c is reached, after which it rises very rapidly. The other term, $\exp(-\Delta G_m/kT)$, is essentially the atomic mobility. Since ΔG_m is constant this decreases rapidly with decreasing temperature. The combination of these terms, i.e. the homogeneous nucleation rate is shown in Fig. 5.4d. Note that at undercoolings smaller than ΔT_c , N is negligible because the driving force ΔG_v is too small, whereas at very high undercoolings N is negligible because diffusion is too slow. A maximum nucleation rate is obtained at intermediate undercoolings. For alloys containing less solute the critical supercooling will not be reached until lower absolute temperatures where diffusion is slower. The resultant variation of N with T in these alloys will therefore appear as shown in Fig. 5.5.

In the above treatment of nucleation it has been assumed that the nucleation rate is constant. In practice however the nucleation rate will initially be low, then gradually rise, and finally decrease again as the first nuclei to form start growing and thereby reduce the supersaturation of the remaining α .

It has also been assumed that the nuclei are spherical with the equilibrium composition and structure of the β phase. However, in practice nucleation will be dominated by whatever nucleus has the minimum activation energy barrier ΔG^* . Equation 5.8 shows that by far the most effective way of minimizing ΔG^* is by the formation of nuclei with the smallest total interfacial energy. In fact this criterion is dominating in nucleation processes. Incoherent nuclei have such a high value of γ that incoherent homogeneous nucleation is virtually impossible. If, however, the nucleus has an *orientation relationship* with the matrix, and coherent interfaces are formed, ΔG^* is greatly reduced

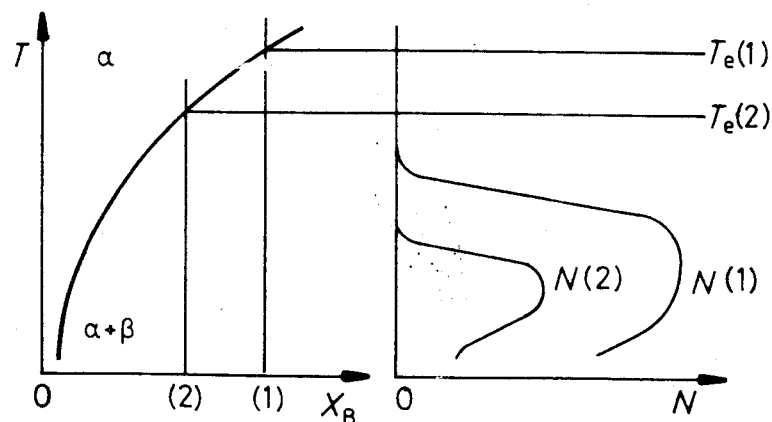


Fig. 5.5 The effect of alloy composition on the nucleation rate. The nucleation rate in alloy 2 is always less than in alloy 1.

and homogeneous nucleation becomes feasible. The formation of a coherent nucleus will of course increase ΔG_s which decreases T'_e . But below T'_e the decrease in γ resulting from coherency can more than compensate for the increase in ΔG_s . Also, by choosing a suitable shape it is often possible to minimize ΔG_s as discussed in Section 3.4.3.

In most systems the α and β phases have such different crystal structures that it is impossible to form coherent low-energy interfaces and homogeneous nucleation of the equilibrium β phase is then impossible. However, it is often possible to form a coherent nucleus of some other, metastable phase (β') which is not present in the equilibrium phase diagram. The most common example of this is the formation of GP zones which will be discussed in more detail later.

There are a few systems in which the equilibrium phase may nucleate homogeneously. For example in the Cu-Co system Cu alloys containing 1–3% Co can be solution treated and quenched to a temperature where Co precipitates. Both Cu and Co are fcc with only a 2% difference in lattice parameter. Therefore very little coherency strain is associated with the formation of coherent Co particles. The interfacial energy is about 200 mJ m^{-2} and the critical undercooling for measurable homogeneous nucleation is about 40°C . This system has been used to experimentally test the theories of homogeneous nucleation and reasonably close agreement was found¹.

Another system in which the equilibrium phase is probably formed homogeneously at a few tens of degrees undercooling is the precipitation of Ni_3Al in many Ni-rich alloys. Depending on the system the misfit varies up to a maximum of 2%, and γ is probably less than 30 mJ m^{-2} . Most other examples of homogeneous nucleation, however, are limited to metastable phases, usually GP zones. (See Section 5.5.1.)

5.2 Heterogeneous Nucleation

Nucleation in solids, as in liquids, is almost always heterogeneous. Suitable nucleation sites are non-equilibrium defects such as excess vacancies, dislocations, grain boundaries, stacking faults, inclusions, and free surfaces, all of which increase the free energy of the material. If the creation of a nucleus results in the destruction of a defect, some free energy (ΔG_d) will be released thereby reducing (or even removing) the activation energy barrier. The equivalent to Equation 5.5 for heterogeneous nucleation is

$$\Delta G_{\text{het}} = -V(\Delta G_v - \Delta G_s) + A\gamma - \Delta G_d \quad (5.18)$$

Nucleation on Grain Boundaries

Ignoring any misfit strain energy, the optimum embryo shape should be that which minimizes the total interfacial free energy. The optimum shape for an *incoherent* grain-boundary nucleus will consequently be two abutted spherical

caps as shown in Fig. 5.6, with θ given by

$$\cos \theta = \gamma_{\alpha\alpha} / 2\gamma_{\alpha\beta} \quad (5.19)$$

(assuming $\gamma_{\alpha\beta}$ is isotropic and equal for both grains). The excess free energy associated with the embryo will be given by

$$\Delta G = -V\Delta G_v + A_{\alpha\beta}\gamma_{\alpha\beta} - A_{\alpha\alpha}\gamma_{\alpha\alpha} \quad (5.20)$$

where V is the volume of the embryo, $A_{\alpha\beta}$ is the area of α/β interface of energy $\gamma_{\alpha\beta}$ created, and $A_{\alpha\alpha}$ the area of α/α grain boundary of energy $\gamma_{\alpha\alpha}$ destroyed during the process. The last term of the above equation is simply ΔG_d in Equation 5.18.

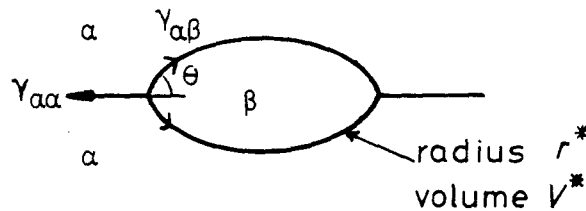


Fig. 5.6 The critical nucleus size (V^*) for grain boundary nucleation.

It can be seen that grain boundary nucleation is analogous to solidification on a substrate (Section 4.1.3) and the same results will apply. Again the critical radius of the spherical caps will be independent of the grain boundary and given by

$$r^* = 2\gamma_{\alpha\beta} / \Delta G_v \quad (5.21)$$

and the activation energy barrier for heterogeneous nucleation will be given by

$$\frac{\Delta G_{het}^*}{\Delta G_{hom}^*} = \frac{V_{het}^*}{V_{hom}^*} = S(\theta) \quad (5.22)$$

where $S(\theta)$ is a shape factor given by

$$S(\theta) = \frac{1}{2}(2 + \cos \theta)(1 - \cos \theta)^2 \quad (5.23)$$

The ability of a grain boundary to reduce ΔG_{het}^* , i.e. its potency as a nucleation site, depends on $\cos \theta$, i.e. on the ratio $\gamma_{\alpha\alpha}/2\gamma_{\alpha\beta}$.

V^* and ΔG^* can be reduced even further by nucleation on a grain edge or grain corner, Figs. 5.7 and 5.8. Figure 5.9 shows how $\Delta G_{het}^*/\Delta G_{hom}^*$ depends on $\cos \theta$ for the various grain boundary nucleation sites.

High-angle grain boundaries are particularly effective nucleation sites for incoherent precipitates with high $\gamma_{\alpha\beta}$. If the matrix and precipitate are sufficiently compatible to allow the formation of lower energy facets then V^* and

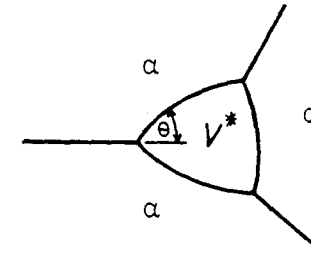


Fig. 5.7 Critical nucleus shape for nucleation on a grain edge.

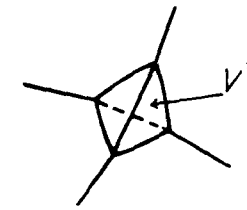


Fig. 5.8 Critical nucleus shape for nucleation on a grain corner.

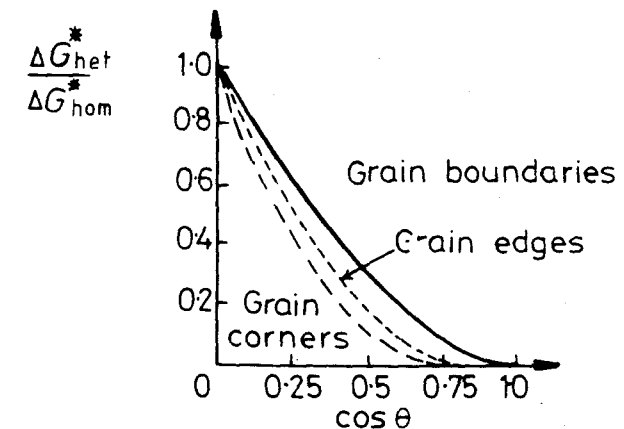


Fig. 5.9 The effect of θ on the activation energy for grain boundary nucleation relative to homogeneous nucleation. (After J.W. Cahn, *Acta Metallurgica* 4 (1956) 449.)

ΔG_{het}^* can be further reduced as shown in Fig. 5.10. The nucleus will then have an orientation relationship with one of the grains. Such nuclei are to be expected whenever possible, since the most successful nuclei, i.e. those which form most rapidly, will have the smallest nucleation barrier.

Other planar defects such as inclusion/matrix interfaces, stacking faults and free surfaces can behave in a similar way to grain boundaries in reducing ΔG^* . Note, however, that stacking faults are much less potent sites due to their lower energy in comparison to high-angle boundaries.

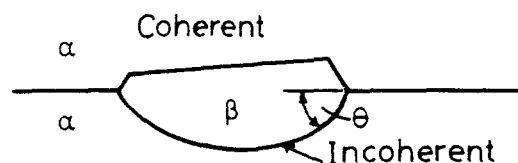


Fig. 5.10 The critical nucleus size can be reduced even further by forming a low-energy coherent interface with one grain.

Dislocations

The lattice distortion in the vicinity of a dislocation can assist nucleation in several ways. The main effect of dislocations is to reduce the ΔG_s -contribution to ΔG^* by reducing the total strain energy of the embryo. A coherent nucleus with a negative misfit, i.e. a smaller volume than the matrix, can reduce its ΔG^* by forming in the region of compressive strain above an edge dislocation, whereas if the misfit is positive it is energetically favourable for it to form below the dislocation.

Nucleation on dislocations may also be assisted by solute segregation which can raise the composition of the matrix to nearer that of the precipitate. The dislocation can also assist in growth of an embryo beyond the critical size by providing a diffusion pipe with a lower ΔG_m .

Dislocations are not very effective for reducing the interfacial energy contribution to ΔG^* . This means that nucleation on dislocations usually requires rather good matching between precipitate and matrix on at least one plane, so that low-energy coherent or semicoherent interfaces can form. Ignoring strain energy effects, the minimum ΔG^* is then achieved when the nucleus shape is the equilibrium shape given by the Wulff construction. When the precipitate and matrix have different crystal structures the critical nucleus should therefore be disc-like or needle-like as discussed in Section 3.4.2.

In fcc crystals the $\frac{a}{2}\langle 110 \rangle$ unit dislocations can dissociate to produce a ribbon of stacking fault, e.g.

$$\frac{a}{2}[110] \rightarrow \frac{a}{6}[121] + \frac{a}{6}[2\bar{1}\bar{1}]$$

giving a stacking fault on $(1\bar{1}1)$ separated by two Shockley partials. Since the stacking fault is in effect four close-packed layers of hcp crystal (Fig 3.59b) it can act as a very potent nucleation site for an hcp precipitate. This type of nucleation has been observed for the precipitation of the hexagonal transition phase γ' in Al-Ag alloys. Nucleation is achieved simply by the diffusion of silver atoms to the fault. Thus there will automatically be an orientation relationship between the γ' precipitate (fault) and the matrix of the type

$$(0001)_{\gamma'} // (1\bar{1}1)_{\alpha}$$

$$[11\bar{2}0]_{\gamma'} // [110]_{\alpha}$$

which ensures good matching and low energy interfaces.

It should be noted that even in annealed specimens dislocation densities are often sufficiently high to account for any precipitate dispersion that is resolvable in the light microscope, i.e. $\sim 1 \mu\text{m}^{-2}$. Figure 5.11 shows an example of niobium carbonitride precipitates on dislocations in a ferritic iron matrix. This is a so-called dark-field electron microscope micrograph in which the precipitates are imaged bright and the matrix dark. The precipitates lie in rows along dislocations.

Excess Vacancies

When an age-hardening alloy is quenched from a high temperature, excess vacancies are retained during the quench. These vacancies can assist nucleation by increasing diffusion rates, or by relieving misfit strain energies. They may influence nucleation either individually or collectively by grouping into small clusters.

Since ΔG_d is relatively small for vacancies, nucleation will only take place when a reasonable combination of the following conditions is met: low interfacial energy (i.e. fully coherent nuclei), small volume strain energy, and

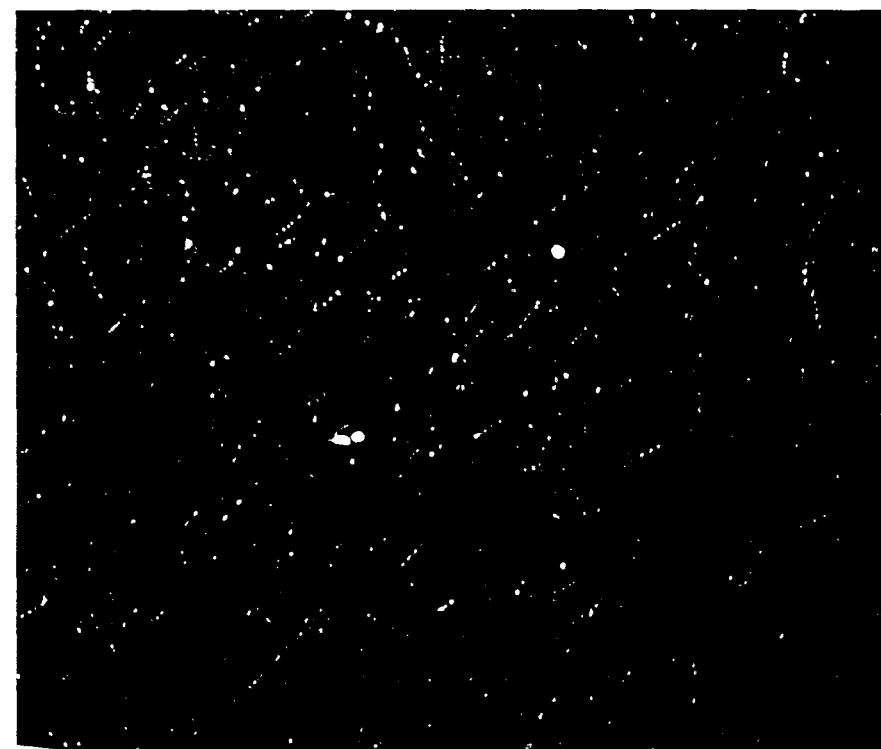


Fig. 5.11 Rows of niobium carbonitride precipitates on dislocations in ferrite ($\times 108\,000$). (Dark-field electron micrograph in which the precipitates show up bright.)

high driving force. These are essentially the same conditions that must be fulfilled for homogeneous nucleation. Since individual vacancies or small clusters cannot be resolved with conventional transmission electron microscopy, evidence for the role of vacancies as heterogeneous nucleation sites is indirect (discussed later).

5.2.1 Rate of Heterogeneous Nucleation

If the various nucleation sites are arranged in order of increasing ΔG_d , i.e. decreasing ΔG^* , the sequence would be roughly

1. homogeneous sites
2. vacancies
3. dislocations
4. stacking faults
5. grain boundaries and interphase boundaries
6. free surfaces.

Nucleation should always occur most rapidly on sites near the bottom of the list. However the relative importance of these sites in determining the overall rate at which the alloy will transform also depends on the relative concentrations of the sites. For homogeneous nucleation every atom is a potential nucleation site, whereas only those atoms on grain boundaries, for example, can take part in boundary-assisted nucleation.

If the concentration of heterogeneous nucleation sites is C_1 per unit volume, the heterogeneous nucleation rate will be given by an equation of the form

$$N_{\text{het}} = \omega C_1 \exp\left(-\frac{\Delta G_m}{kT}\right) \cdot \exp\left(-\frac{\Delta G^*}{kT}\right) \text{ nuclei m}^{-3}\text{s}^{-1} \quad (5.24)$$

This is plotted as a function of temperature in Fig. 5.12. Note that, as with heterogeneous nucleation in liquids, measurably high nucleation rates can be

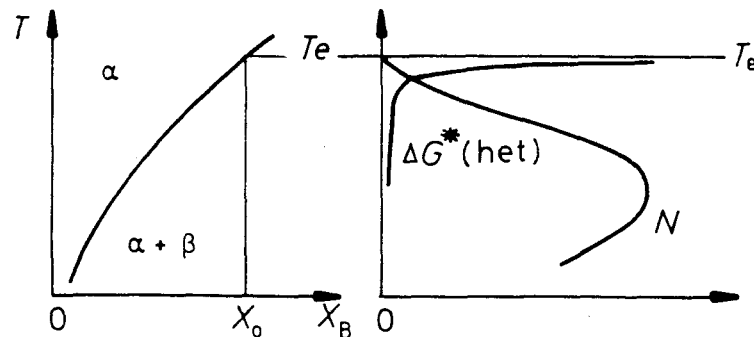


Fig. 5.12 The rate of heterogeneous nucleation during precipitation of β in alloy X_0 as a function of undercooling.

obtained at very small driving forces. The relative magnitudes of the heterogeneous and homogeneous volume nucleation rates can be obtained by dividing Equation 5.11 by 5.24 giving

$$\frac{N_{\text{het}}}{N_{\text{hom}}} = \frac{C_1}{C_0} \exp\left(\frac{\Delta G_{\text{hom}}^* - \Delta G_{\text{het}}^*}{kT}\right) \quad (5.25)$$

(Differences in ω and ΔG_m are not so important and have been ignored.) Since ΔG^* is always smallest for heterogeneous nucleation the exponential factor in the above equation is always a large quantity which favours a high heterogeneous nucleation rate. However, the factor (C_1/C_0) must also be taken into account, i.e. the number of atoms on heterogeneous sites relative to the number within the matrix. For grain boundary nucleation

$$\frac{C_1}{C_0} = \frac{\delta}{D} \quad (5.26)$$

where δ is the boundary thickness and D is the grain size. For nucleation on grain edges and corners (C_1/C_0) becomes reduced even further to $(\delta/D)^2$ and $(\delta/D)^3$. Therefore for a 50 μm grain size taking δ as 0.5 nm gives $\delta/D \approx 10^{-5}$. Consequently grain boundary nucleation will dominate over homogeneous nucleation if the boundary is sufficiently potent to make the exponential term in Equation 5.23 greater than 10^5 . Values for C_1/C_0 for other sites are listed in Table 5.1.

In general the type of site which gives the highest volume nucleation rate will depend on the driving force (ΔG_v). At very small driving forces, when activation energy barriers for nucleation are high, the highest nucleation rates will be produced by grain-corner nucleation. As the driving force increases, however, grain edges and then boundaries will dominate the transformation. At very high driving forces it may be possible for the (C_1/C_0) term to dominate and then homogeneous nucleation provides the highest nucleation rates. Similar considerations will apply to the relative importance of other heterogeneous nucleation sites.

The above comments concerned nucleation during isothermal transformations when the specimen is held at a constant temperature. If nucleation occurs during continuous cooling the driving force for nucleation will increase with time. Under these conditions the initial stages of the transformation will be dominated by those nucleation sites which can first produce a measurable volume nucleation rate. Considering only grain boundaries again, if $\gamma_{\alpha\alpha}/\gamma_{\alpha\beta}$ is high, noticeable transformation will begin first at the grain corners, whereas if the grain boundary is less potent ($\gamma_{\alpha\alpha}/\gamma_{\alpha\beta}$ smaller) nucleation may not be possible until such high driving forces are reached that less favourable heterogeneous or even homogeneous nucleation sites dominate. This will not of course exclude precipitation on potent heterogeneous sites, but they will make only a very small contribution to the total nucleation rate.

Table 5.1 C_1/C_0 for Various Heterogeneous Nucleation Sites

Grain boundary	Grain edge	Grain corner	Dislocations	Excess vacancies
$D = 50 \mu\text{m}$	$D = 50 \mu\text{m}$	$D = 50 \mu\text{m}$	10^5 mm^{-2}	$X_v = 10^{-6}$
10^{-5}	10^{-10}	10^{-15}	10^{-8}	10^{-6}
			10^8 mm^{-2}	
			10^{-5}	

5.3 Precipitate Growth

As explained above, the successful critical nuclei are those with the smallest nucleation barrier, i.e. the smallest critical volume. In the absence of strain-energy effects the precipitate shape satisfying this criterion is that which minimizes the total interfacial free energy. Thus nuclei will usually be bounded by a combination of coherent or semicoherent facets and smoothly curved incoherent interfaces. For the precipitate to grow these interfaces must migrate and the shape that develops during growth will be determined by the relative migration rates. As explained in Section 3.5.1, when the two phases have different crystal structures semicoherent interfaces have very low mobility and are forced to migrate by a ledge mechanism. Incoherent interfaces on the other hand are highly mobile. If there are problems in maintaining a constant supply of ledges the incoherent interfaces will be able to advance faster than the semicoherent interface and a nucleus with one plane of good matching should grow into a thin disc or plate as shown in Fig. 5.13. This is the origin of the so-called Widmanstätten morphology¹².

The next few sections will be concerned with developing an approximate quantitative treatment for the ledge mechanism and for the rate of migration of curved incoherent interfaces, but before treating these two cases it is useful to begin with the simpler case of a planar incoherent interface.

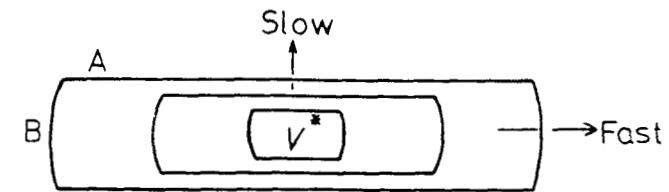


Fig. 5.13 The effect of interface type on the morphology of a growing precipitate. (A) Low-mobility semicoherent interfaces. (B) High-mobility incoherent interfaces.

5.3.1 Growth behind Planar Incoherent Interfaces

It will be apparent from the above discussion that planar interfaces in crystalline solids will usually not be incoherent. However, one situation where approximately planar incoherent interfaces may be found is after grain-boundary nucleation. If many incoherent nuclei form on a grain boundary they might subsequently grow together to form a slab of β precipitate as shown in Fig. 5.14.

Imagine that such a slab of solute-rich precipitate has grown from zero thickness and that the instantaneous growth rate is v . Since the concentration of solute in the precipitate (C_β) is higher than in the bulk (C_0) the matrix adjacent to the precipitate will be depleted of solute as shown. Also since the interface is incoherent diffusion-controlled growth and local equilibrium at

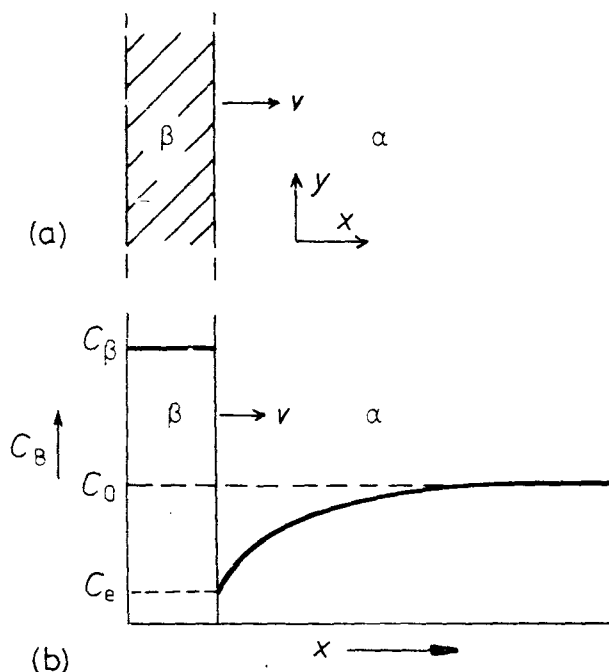


Fig. 5.14 Diffusion-controlled thickening of a precipitate plate.

the interface can be assumed, i.e. the solute concentration in the matrix adjacent to the β will be the equilibrium value C_e . The growth rate (v) will depend on the concentration gradient at the interface dC/dx .

For unit area of interface to advance a distance dx a volume of material $1 \cdot dx$ must be converted from α containing C_e to β containing C_β moles of B per unit volume, i.e. $(C_\beta - C_e)dx$ moles of B must be supplied by diffusion through the α . The flux of B through unit area in time dt is given by $D(dC/dx)dt$, where D is the *interdiffusion* coefficient (or interstitial diffusion coefficient). Equating these two quantities gives

$$v = \frac{dx}{dt} = \frac{D}{C_\beta - C_e} \cdot \frac{dC}{dx} \quad (5.27)$$

As the precipitate grows solute must be depleted from an ever-increasing volume of matrix so that dC/dx in the above equation decreases with time. To make this quantitative, consider a simplified approach originally due to Zener². If the concentration profile is simplified to that shown in Fig. 5.15 dC/dx is given by $\Delta C_0/L$ where $\Delta C_0 = C_0 - C_e$. The width of the diffusion zone L can be obtained by noting that the conservation of solute requires the two shaded areas in Fig. 5.15 to be equal, i.e.

$$(C_\beta - C_0)x = L\Delta C_0/2$$

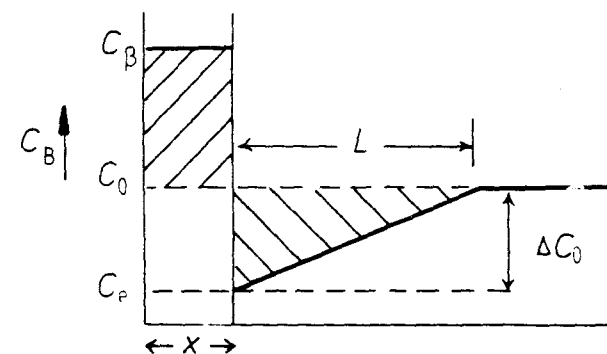


Fig. 5.15 A simplification of the concentration profile.

where x is the thickness of the slab. The growth rate therefore becomes

$$v = \frac{D(\Delta C_0)^2}{2(C_\beta - C_e)(C_\beta - C_0)x} \quad (5.28)$$

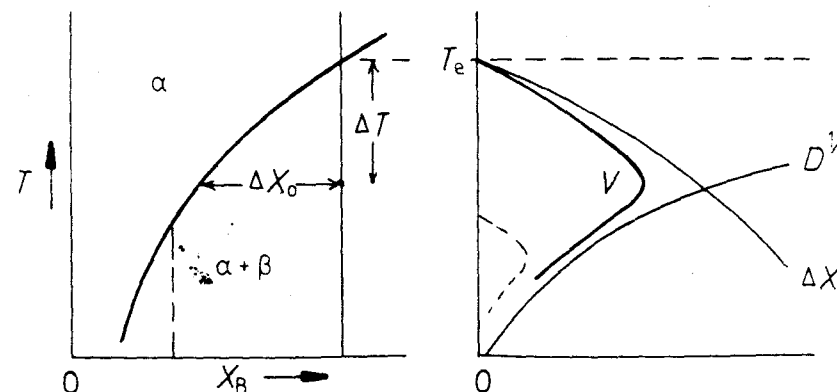
If it is assumed that the molar volume (V_m) is a constant, the concentrations in the above equation can be replaced by mole fractions ($X = CV_m$). Furthermore, for the sake of simplicity it can often be assumed that $C_\beta - C_0 \approx C_\beta - C_e$. Integration of Equation 5.28 then gives

$$x = \frac{\Delta X_0}{\sqrt{(X_\beta - X_e)}} \sqrt{Dt} \quad (5.29)$$

and

$$v = \frac{\Delta X_0}{2(X_\beta - X_e)} \sqrt{\frac{D}{t}} \quad (5.30)$$

where $\Delta X_0 = X_0 - X_e$ (Fig. 5.16) is the supersaturation prior to precipitation.

Fig. 5.16 The effect of temperature and position on growth rate, v .

The following points are important to note regarding these equations.

1. $x \propto \sqrt{(Dt)}$, i.e. precipitate thickening obeys a parabolic growth law.
2. $v \propto \Delta X_0$, i.e. for a given time the growth rate is proportional to the supersaturation.
3. $v \propto \sqrt{(D/t)}$.

The effect of alloy composition and temperature on growth rate is illustrated in Fig. 5.16. Growth rates are low at small undercoolings due to small supersaturation ΔX_0 but are also low at large undercoolings due to slow diffusion. A maximum growth rate will occur at some intermediate undercooling.

When the diffusion fields of separate precipitates begin to overlap Equation 5.30 will no longer apply, but growth will decelerate more rapidly and finally cease when the matrix concentration is X_e everywhere. Fig. 5.17.

Although these equations are only approximate and were derived for a planar interface, the conclusions are not significantly altered by more thorough treatments or by allowing curved interfaces. Thus it can be shown that

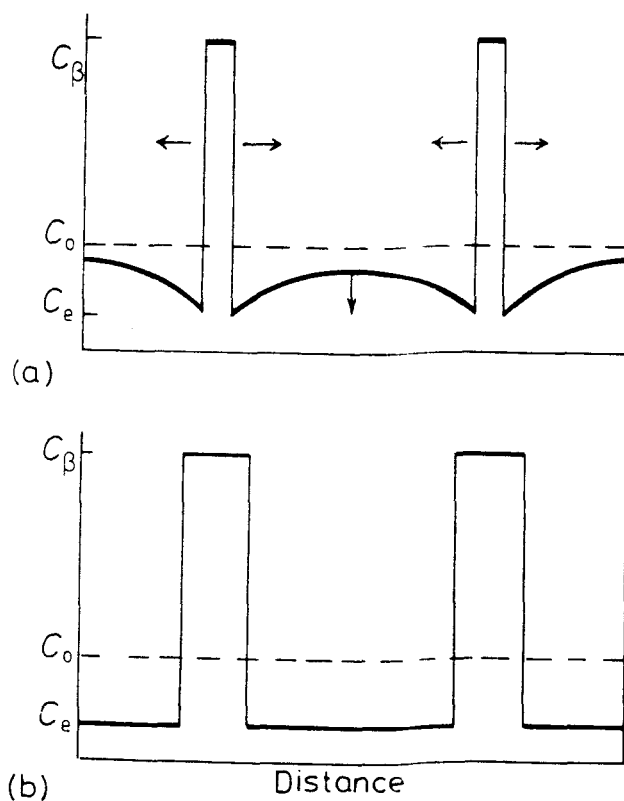


Fig. 5.17 (a) Interference of growing precipitates due to overlapping diffusion fields at later stage of growth. (b) Precipitate has stopped growing.

any linear dimension of a spheroidal precipitate increases as $\sqrt{(Dt)}$ provided all interfaces migrate under volume diffusion control.

Usually grain boundary precipitates do not form a continuous layer along the boundary but remain as isolated particles. The growth of such precipitates can occur at rates far greater than allowed by volume diffusion. The reason for this is that the grain boundary can act as a collector plate for solute as shown in Fig. 5.18.²² Growth of such a so-called grain-boundary allotriomorph involves three steps: (1) volume diffusion of solute to the grain boundary; (2) diffusion of solute along the grain boundary with some attachment at the precipitate rim; and (3) diffusion along the α/β interfaces allowing accelerated thickening. This mechanism is of greatest significance when substitutional diffusion is involved. In the case of interstitial solutions diffusion short circuits are comparatively unimportant due to the high volume diffusion rates.

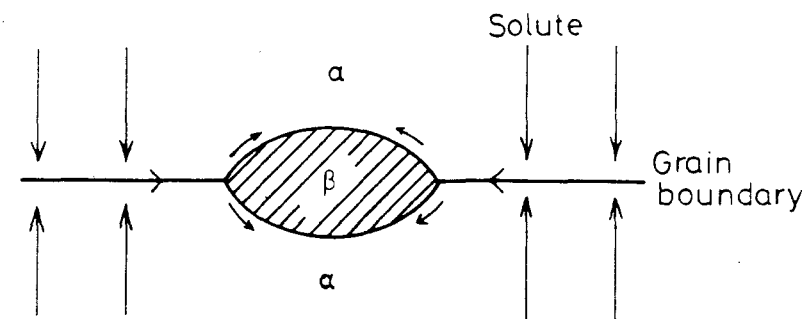


Fig. 5.18 Grain-boundary diffusion can lead to rapid lengthening and thickening of grain boundary precipitates.

5.3.2 Diffusion-Controlled Lengthening of Plates or Needles

Imagine now that the β precipitate is a plate of constant thickness having a cylindrically curved *incoherent* edge of radius r as shown in Fig. 5.19a. Again the concentration profile across the curved interface will appear as shown in Fig. 5.19b, but now, due to the Gibbs-Thomson effect, the equilibrium concentration in the matrix adjacent to the edge will be increased to C_r . The concentration gradient available to drive diffusion to the advancing edge is therefore reduced to $\Delta C/L$ where $\Delta C = C_0 - C_r$ and L is a characteristic diffusion distance. The diffusion problem in this case is more complex as diffusion occurs radially. However, solution of the relevant equations shows that L is given by kr where k is a numerical constant (~ 1). By analogy with Equation 5.27, therefore, the lengthening rate will be given by

$$v = \frac{D}{C_\beta - C_r} \cdot \frac{\Delta C}{kr} \quad (5.31)$$

The composition difference available to drive diffusion will depend on the tip radius as shown in Fig. 5.20. With certain simplifying assumptions it can

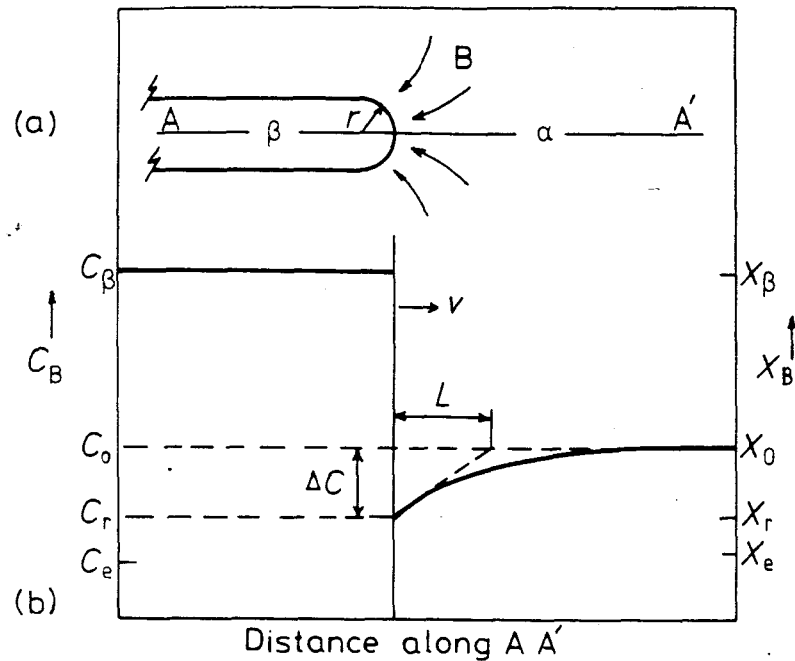


Fig. 5.19 (a) The edge of a plate-like precipitate. (b) A concentration profile along AA' in (a).

be shown that

$$\Delta X = \Delta X_0 \left(1 - \frac{r^*}{r} \right) \quad (5.32)$$

where $\Delta X = X_0 - X_r$, $\Delta X_0 = X_0 - X_e$ and r^* is the critical nucleus radius, i.e. the value of r required to reduce ΔX to zero. Again, assuming constant molar volume, the above equations can be combined to give

$$v = \frac{D \Delta X_0}{k(X_\beta - X_r)} \cdot \frac{1}{r} \left(1 - \frac{r^*}{r} \right) \quad (5.33)$$

This equation will apply as long as there is no decrease in supersaturation far from the interface due to other precipitates. The difference between this equation and Equation 5.30 is that for a given plate thickness the lengthening rate should be constant, i.e. $x \propto t$ (linear growth).

Although the above equations were developed for the lengthening of a plate, the same equations can be derived for the lengthening of a needle under diffusion-controlled growth. The only difference is that the edge of a needle has a spherical tip so that the Gibbs-Thomson increase in free energy is $2\gamma V_m/r$ instead of $\gamma V_m/r$. The value of r^* in Equation 5.33 will, therefore, be different for a plate and a needle.

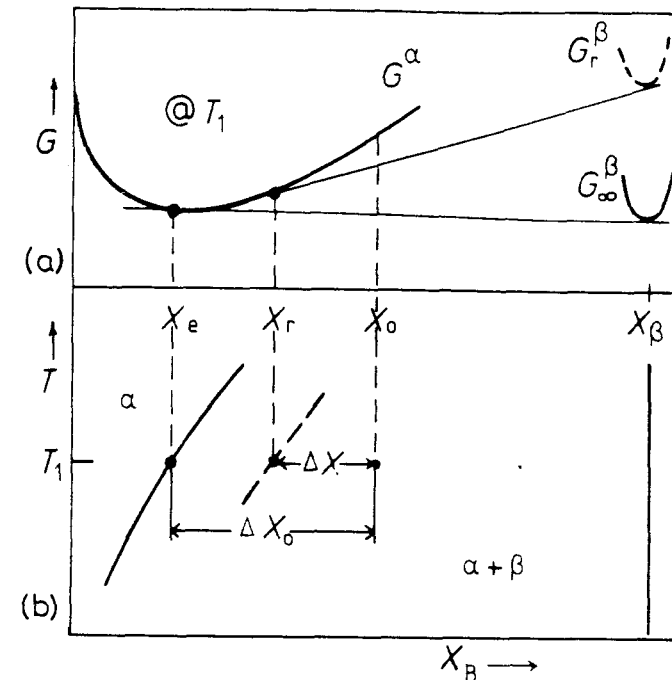


Fig. 5.20 The Gibbs-Thomson effect. (a) Free energy curves at T_1 . (b) Corresponding phase diagram.

The above treatment only applies to plates or needles that lengthen by a volume diffusion-controlled continuous growth process. This is a reasonable assumption for the curved ends of needles, but in the case of plate-like precipitates the edges are often faceted and are observed to migrate by a ledge mechanism. Atoms can then only attach at the ledges and new equations must be derived as discussed below.

Another source of deviation between theory and practice is if solute can be transported to the advancing precipitate edges by short-circuit diffusion in the broad faces of the precipitate plate.

5.3.3 Thickening of Plate-like Precipitates

The treatment given in Section 5.3.1 for a planar incoherent interface is only valid for interfaces with high accommodation factors. In general this will not be the case for the broad faces of plate-like precipitates which are semicoherent and are restricted to migrate by the lateral movement of ledges.

For simplicity, imagine a plate-like precipitate that is thickening by the lateral movement of linear ledges of constant spacing λ and height h , Fig. 5.21. It can readily be seen that the half-thickness of the plate should

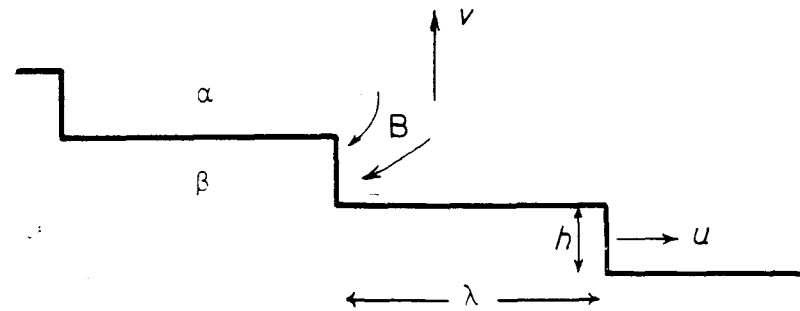


Fig. 5.21 Thickening of plate-like precipitates by ledge mechanism.

increase at a rate v given by

$$v = \frac{uh}{\lambda} \quad (5.34)$$

where u is the rate of lateral migration.

The problem of ledge migration is very similar to that of plate lengthening. The necessary composition changes required for precipitate growth must be achieved by long-range diffusion to and from the ledges as shown in Fig. 5.21. If the edges of the ledges are incoherent the matrix composition in contact with the ledges will be X_e and growth will be diffusion controlled. A similar treatment to that given in Section 5.3.2 then gives the rate of lateral migration as³

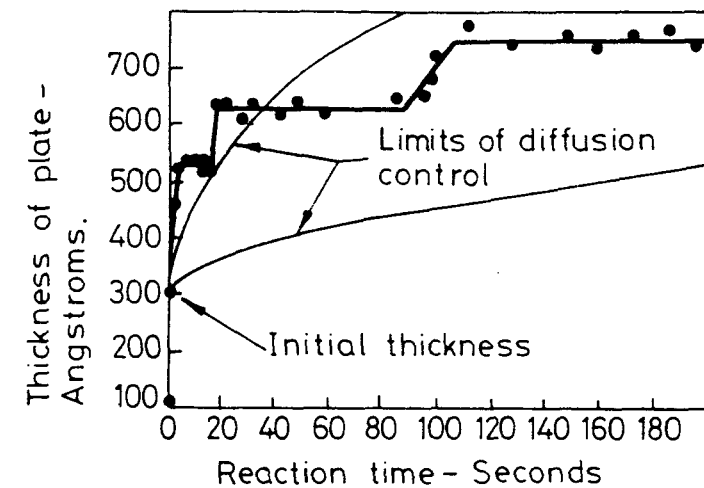
$$u = \frac{D\Delta X_0}{k(X_\beta - X_e)h} \quad (5.35)$$

This is essentially the same as Equation 5.33 for the lengthening of a plate with $h \approx r$ and $X_r = X_e$, i.e. no Gibbs-Thomson effect. Combining the above equations shows that the thickening rate is independent of h and given by

$$v = \frac{D\Delta X_0}{k(X_\beta - X_e)\lambda} \quad (5.36)$$

Thus, provided the diffusion fields of different precipitates do not overlap, the rate at which plates thicken will be inversely proportional to the interledge spacing λ . The validity of Equation 5.36 is dependent on there being a constant supply of ledges. As with faceted solid/liquid interfaces, new ledges can be generated by various mechanisms such as repeated surface nucleation, spiral growth, nucleation at the precipitate edges, or from intersections with other precipitates. With the exception of spiral growth, however, none of these mechanisms can maintain a supply of ledges with constant λ .

By using hot-stage transmission electron microscopy it is possible to measure the thickening rates of individual precipitate plates. Figure 5.22 shows results obtained from a γ plate in the Al-Ag system⁴. It can be seen that there

Fig. 5.22 The thickening of a γ plate in an Al-15 wt% Ag alloy at 400 °C. (From C. Laird and H.I. Aaronson, *Acta Metallurgica* 17 (1969) 505.)

are appreciable intervals of time when there is no perceptible increase in plate thickness followed by periods when the thickness increases rapidly as an interfacial ledge passes. The two smooth lines in the figure are upper and lower limits for the rate of thickening for a planar incoherent interface in the same system, assuming diffusion control. The ledge mechanism is clearly a very different process. The fact that there is no perceptible increase in thickness except when ledges pass is strong evidence in favour of the immobility of semicoherent interfaces. It can also be seen that the thickening rate is not constant implying that ledge nucleation is rate controlling.

Measurements on precipitates in other systems indicate that even within the same system the thickness/time relationship can vary greatly from plate to plate, presumably depending on differences in the ease of nucleation of new ledges.

5.4 Overall Transformation Kinetics—TTT Diagrams

The progress of an isothermal phase transformation can be conveniently represented by plotting the fraction transformation (f) as a function of time and temperature, i.e. a TTT diagram as shown in Fig. 5.23a for example. For transformations of the type $\alpha \rightarrow \beta$, f is just the volume fraction of β at any time. For precipitation reactions $\alpha' \rightarrow \alpha + \beta$, f can be defined as the volume of β at time t divided by the final volume of β . In both cases f varies from 0 to 1 from the beginning to the end of the transformation, Fig. 5.23b.

Among the factors that determine $f(t, T)$ are the nucleation rate, the growth rate, the density and distribution of nucleation sites, the overlap of diffusion fields from adjacent transformed volumes, and the impingement of

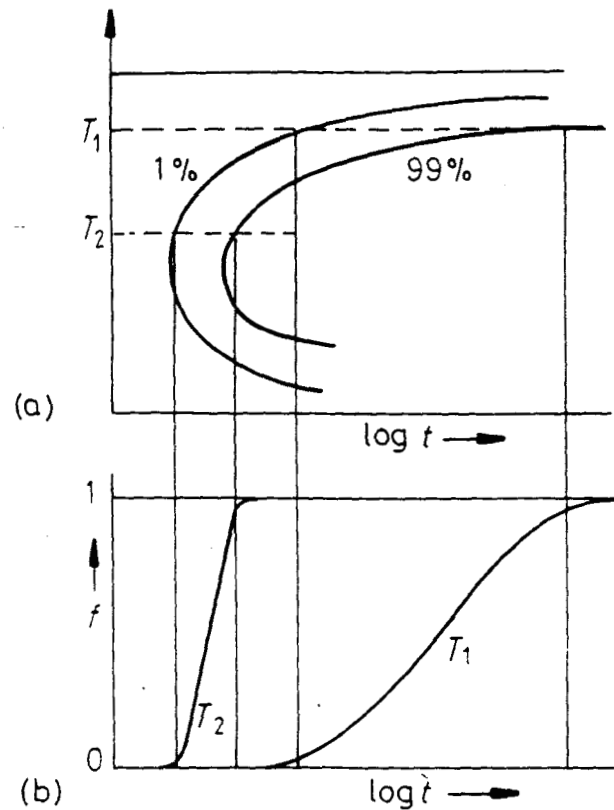


Fig. 5.23 The percentage transformation versus time for different transformation temperatures.

adjacent transformed volumes. Some of the problems involved are illustrated in Fig. 5.24. After quenching to the transformation temperature the metastable α phase will contain many nucleation sites (usually heterogeneous). One possible sequence of events, Fig. 5.24a, is that nuclei form throughout the transformation so that a wide range of particle sizes exists at any time. Another possibility is that all nuclei form right at the beginning of transformation, Fig. 5.24b. If all potential nucleation sites are consumed in the process this is known as *site saturation*. In Fig. 5.24a, f will depend on the nucleation rate and the growth rate. In Fig. 5.24b, f will only depend on the number of nucleation sites and the growth rate. For transformations of the type $\alpha \rightarrow \beta$ or $\alpha \rightarrow \beta + \gamma$ (known collectively as *cellular transformations*) all of the parent phase is consumed by the transformation product, Fig. 5.24c. In these cases the transformation does not terminate by the gradual reduction in the growth rate, but by the impingement of adjacent cells growing with a constant velocity. Pearlite, cellular precipitation, massive transformations and recrystallization belong to this category.

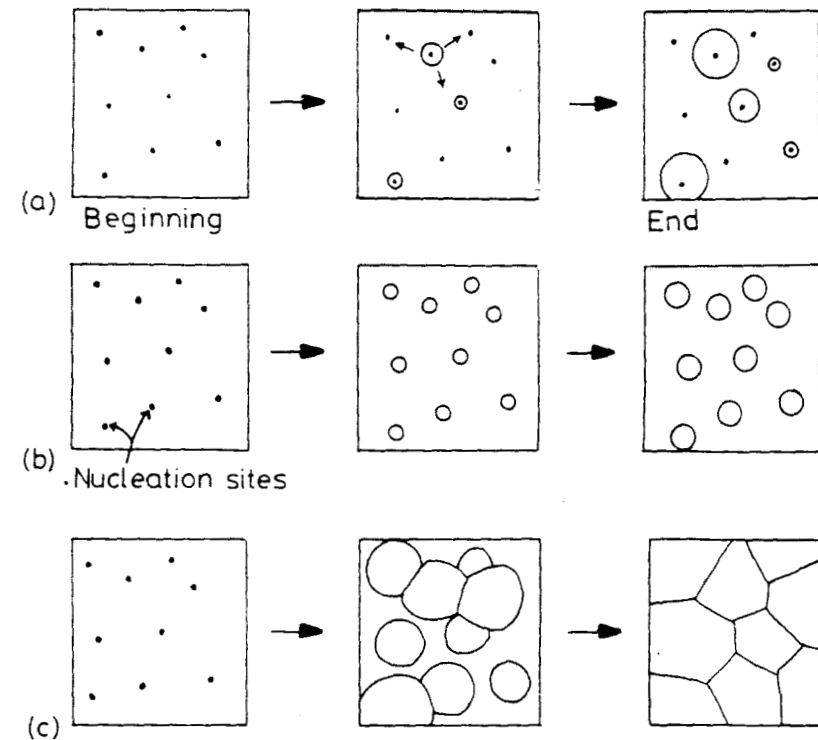


Fig. 5.24 (a) Nucleation at a constant rate during the whole transformation. (b) Site saturation—all nucleation occurs at the beginning of transformation. (c) A cellular transformation.

As a simple example of the derivation of $f(t, T)$ consider a cellular transformation ($\alpha \rightarrow \beta$) in which β cells are continuously nucleated throughout the transformation at a constant rate N . If the cells grow as spheres at a constant rate v , the volume of a cell nucleated at time zero will be given by

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(vt)^3$$

A cell which does not nucleate until time τ will have a volume

$$V' = \frac{4}{3}\pi v^3(t - \tau)^3$$

The number of nuclei that formed in a time increment of $d\tau$ will be $Nd\tau$ per unit volume of untransformed α . Thus if the particles do not impinge on one another, for a unit total volume

$$f = \Sigma V' = \frac{4}{3}\pi N v^3 \int_0^t (t - \tau)^3 d\tau$$

i.e.

$$f = \frac{\pi}{3} N v^3 t^4 \quad (5.37)$$

This equation will only be valid for $f \ll 1$. As time passes the β cells will eventually impinge on one another and the rate of transformation will decrease again. The equation valid for randomly distributed nuclei for both long and short times is⁶

$$f = 1 - \exp\left(-\frac{\pi}{3} N v^3 t^4\right) \quad (5.38)$$

Note that this is the same as Equation 5.37 for short times, since $1 - \exp(-z) \approx z$ when $z \ll 1$. It is also reasonable for long times since as $t \rightarrow \infty$, $f \rightarrow 1$.

Equation 5.38 is known as a *Johnson-Mehl-Avrami equation*. In general, depending on the assumptions made regarding the nucleation and growth processes, a variety of similar equations can be obtained with the form

$$f = 1 - \exp(-kt^n) \quad (5.39)$$

where n is a numerical exponent whose value can vary from ~ 1 to 4. Provided there is no change in the nucleation mechanism, n is independent of temperature. k , on the other hand, depends on the nucleation and growth rates and is therefore very sensitive to temperature. For example, in the case above, $k = \pi N v^3 / 3$ and both N and v are very temperature sensitive.

Since $\exp(-0.7) = 0.5$ the time for 50% transformation ($t_{0.5}$) is given by

$$kt_{0.5}^n = 0.7$$

i.e.

$$t_{0.5} = \frac{0.7^{1/n}}{k^{1/n}} \quad (5.40)$$

For the case discussed above

$$t_{0.5} = \frac{0.9}{N^{1/4} v^{3/4}} \quad (5.41)$$

Consequently it can be seen that rapid transformations are associated with large values of k , i.e. rapid nucleation and growth rates, as expected.

Civilian transformations that occur on cooling are typified by C-shaped TTT curves as shown in Fig. 5.23a. This can be explained on the basis of the variation of nucleation and growth rates with increasing undercooling. At temperatures close to T_e the driving force for transformation is very small so that both nucleation and subsequent growth rates are slow and a long time is required for transformation. When ΔT is very large, on the other hand, slow diffusion rates limit the rate of transformation. A maximum rate is, therefore, obtained at intermediate temperatures.

5.5 Precipitation in Age-Hardening Alloys

The theory of nucleation and growth that has been described above is able to provide general guidelines for understanding civilian transformations. Let us now turn to a consideration of some examples of the great variety of civilian transformations that can occur in solids, and begin with alloys that can be age-hardened. These alloys are characterized by phase diagrams such as that shown in Fig. 5.1a(i). Two extensively researched and illustrative examples are aluminium-copper and aluminium-silver alloys.

5.5.1 Precipitation in Aluminium-Copper Alloys

GP Zones

Figure 5.25 shows the Al-rich end of the Al-Cu phase diagram. If an alloy with the composition Al-4 wt% Cu (1.7 atomic %) is heated to a temperature of about 540 °C all copper will be in solid solution as a stable fcc α phase, and by quenching the specimen rapidly into water there is no time for any transformation to occur so that the solid solution is retained largely unchanged to room temperature. However, the solid solution is now supersaturated with Cu and there is a driving force for precipitation of the equilibrium θ phase, CuAl_2 .

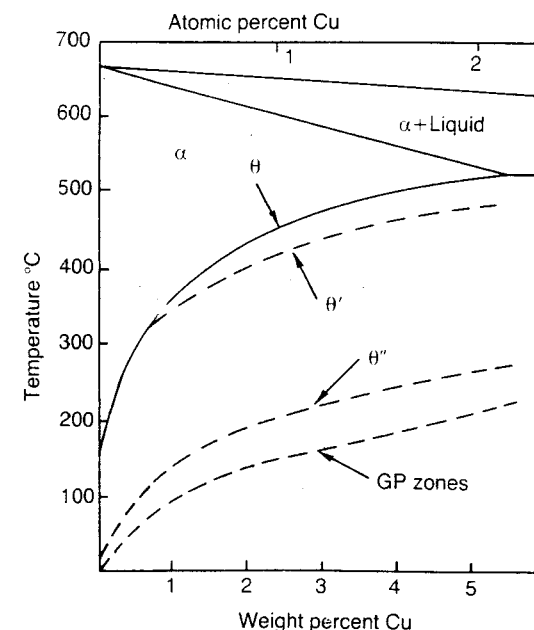


Fig. 5.25 Al-Cu phase diagram showing the metastable GP zone, θ'' and θ' solvuses. (Reproduced from G. Lorimer, *Precipitation Processes in Solids*, K.C. Russell and H.I. Aaronson (Eds.), The Metallurgical Society of AMIE, 1978, p. 87.)

If the alloy is now *aged* by holding for a period of time at room temperature or some other temperature below about 130 °C it is found that the first precipitate to nucleate is not θ but coherent Cu-rich GP zones. (Copper-rich zones in Al-Cu alloys were detected independently in 1938 by Guinier and Preston from streaks in X-ray diffraction patterns.) The reason for this can be understood on the basis of the relative activation energy barriers for nucleation as discussed earlier. GP zones are fully coherent with the matrix and therefore have a very low interfacial energy, whereas the θ phase has a complex tetragonal crystal structure which can only form with high-energy incoherent interfaces. In addition, the zones minimize their strain energy by choosing a disc-shape perpendicular to the elastically soft $\langle 100 \rangle$ directions in the fcc matrix, Fig. 5.26. Therefore, despite the fact that the driving force for precipitation of GP zones ($\Delta G_v - \Delta G_s$) is less than for the equilibrium phase, the barrier to nucleation (ΔG^*) is still less, and the zones nucleate most rapidly. The microstructure of an Al-Cu alloy aged to produce GP zones is shown in Fig. 5.30a. These zones are about 2 atomic layers thick and 10 nm in diameter with a spacing of ~ 10 nm. The zones themselves are not resolved. The contrast in the image is due to the coherency misfit strain perpendicular to the zones. This distorts the lattice causing local variations in the intensity of electron diffraction, which in turn shows up as variations in the image intensity. Microstructurally, the zones appear to be homogeneously nucleated, however excess vacancies are thought to play an important role in their formation. This point will be returned to later.

GP zones are formed as the first precipitate during low-temperature ageing of many technologically important alloys, notably those based on aluminium (see Tables 5.2 and 5.3). In dilute Al-Zn and Al-Ag alloys Zn-rich and Ag-rich GP zones are found. In these cases there is very little misfit strain and ΔG^* is minimized by the formation of spherical zones with a minimum interfacial energy, Fig. 3.39.

Transition Phases

The formation of GP zones is usually followed by the precipitation of so-called transition phases. In the case of Al-Cu alloys the equilibrium θ phase is

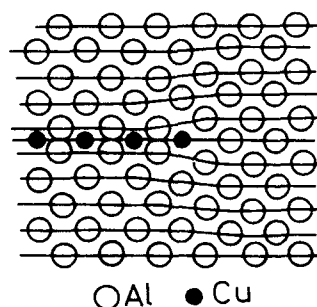


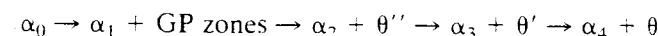
Fig. 5.26 Section through a GP zone parallel to the (200) plane. (Based on the work of V. Gerold: *Zeitschrift für Metallkunde* 45 (1954) 599.)

Table 5.2 Some Precipitation-Hardening Sequences

(Mainly from J.W. Martin, *Precipitation Hardening*, Pergamon Press, Oxford, 1968.)

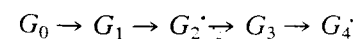
Base metal	Alloy	Precipitation sequence
Aluminium	Al-Ag	GPZ (spheres) $\rightarrow$ γ' (plates) $\rightarrow$ γ (Ag_2Al)
	Al-Cu	GPZ (discs) $\rightarrow$ θ'' (discs) $\rightarrow$ θ' (plates) $\rightarrow$ θ (CuAl_2)
	Al-Cu-Mg	GPZ (rods) $\rightarrow$ S' (laths) $\rightarrow$ S (CuMgAl_2) (laths)
	Al-Zn-Mg	GPZ (spheres) $\rightarrow$ η' (plates) $\rightarrow$ η (MgZn_2) (plates or rods)
Copper	Al-Mg-Si	GPZ (rods) $\rightarrow$ β' (rods) $\rightarrow$ β (Mg_2Si) (plates)
	Cu-Be	GPZ (discs) $\rightarrow$ $\gamma' \rightarrow \gamma$ (CuBe)
Iron	Cu-Co	GPZ (spheres) $\rightarrow$ β (Co) (plates)
	Fe-C	ϵ -carbide (discs) $\rightarrow$ Fe_3C (plates)
Nickel	Fe-N	α'' (discs) $\rightarrow$ Fe_4N
	Ni-Cr-Ti-Al	γ' (cubes or spheres)

preceded by θ'' and θ' . The total precipitation process can be written



where α_0 is the original supersaturated solid solution, α_1 is the composition of the matrix in equilibrium with GP zones, α_2 the composition in equilibrium with θ'' etc.

Figure 5.27 shows a schematic free energy diagram for the above phases. Since GP zones and the matrix have the same crystal structure they lie on the same free energy curve (ignoring strain energy effects—see Section 5.5.5). The transition phases θ'' and θ' are less stable than the equilibrium θ phase and consequently have higher free energies as shown. The compositions of the matrix in equilibrium with each phase— α_1 , α_2 , α_3 , α_4 —are given by the common tangent construction. These compositions correspond to points on the solvus lines for GP zones, θ'' , θ' and θ shown in Fig. 5.25. The free energy of the alloy undergoing the above precipitation sequence decreases as



as shown in Fig. 5.27. *Transformation stops when the minimum free energy equilibrium state G_4 is reached, i.e. $\alpha_4 + \theta$.

Transition phases form because, like GP zones, they have a lower activation energy barrier for nucleation than the equilibrium phase, Fig. 5.28a. The free energy of the alloy therefore decreases more rapidly via the transition phases than by direct transformation to the equilibrium phase, Fig. 5.28b.

Table 5.3 Mechanical Properties of some Commercial Precipitation-Hardening Alloys

Base metal	Alloy	Composition (wt%)	Precipitate	YS* MPa	UTS* MPa	Elongation* %
Aluminium	2024	Cu (4.5) Mg (1.5) Mn (0.6)	S' (Al ₂ CuMg)	390	500	13
	6061	Mg (1.0) Si (0.6) Cu (0.25)	β' (Mg ₂ Si)	280	315	12
	7075	Cr (0.2)	η' (MgZn ₂)	500	570	11
Copper	Cu-Be Nimonic 105	Zn (5.6) Mg (2.5) Cu (1.6)	zones	770	1160	5
		Mn (0.2) Cr (0.3)	γ' (Ni ₃ TiAl)	750†	1100†	25†
Nickel	Maraging Steel	Be (1.9) Co (0.5)	σ (FeMo)	1000	1900	4
		Co (20) Cr (15) Mo (5)	+ Ni ₃ Ti			
Iron		Al (4.5) Ti (1.0) C (0.15)				
		Ni (18) Co (9) Mo (5)				
		Ti (0.7) Al (0.1)				

* At peak hardness tested at room temperature.

† Tested at 600 °C.

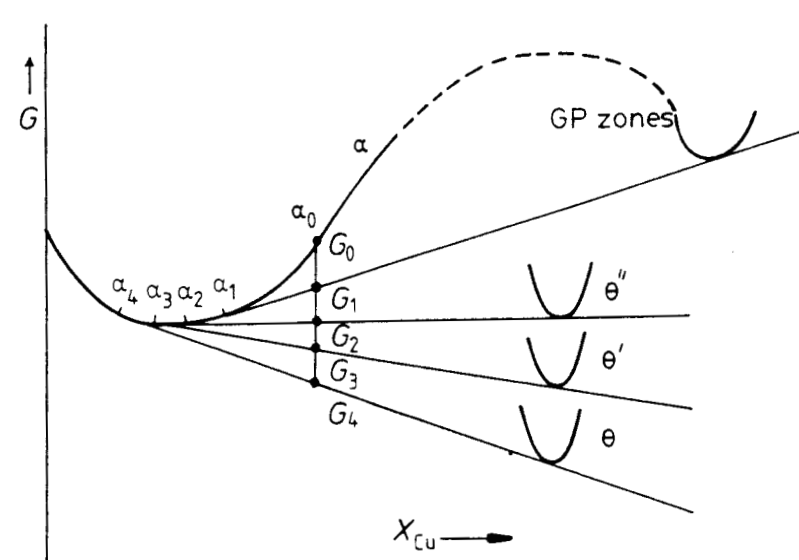


Fig. 5.27 A schematic molar free energy diagram for the Al-Cu system.

The lower activation energy barriers are achieved because the crystal structures of the transition phases are intermediate between those of the matrix and the equilibrium phase. In this way the transition phases can achieve a high degree of coherence and thus a low interfacial energy contribution to ΔG^* . The equilibrium phase on the other hand usually has a complex crystal structure that is incompatible with the matrix and results in high-energy interfaces and high ΔG^* .

The crystal structures of θ'' , θ' and θ are shown in Fig. 5.29 along with that of the fcc matrix for comparison. θ'' has a tetragonal unit cell which is essentially a distorted fcc structure in which the copper and aluminium atoms are ordered on (001) planes as shown. Note that the atomic structure of the (001) planes is identical to that in the matrix, and the (010) and (100) planes are very similar, apart from a small distortion in the [001] direction. θ'' forms as fully coherent plate-like precipitates with a $\{001\}_\alpha$ habit plane and the following orientation relationship to the matrix:

$$\begin{aligned} (001)_{\theta''} &\parallel (001)_\alpha \\ [100]_{\theta''} &\parallel [100]_\alpha \end{aligned}$$

A high magnification transmission electron micrograph of an alloy aged to produce θ'' precipitates is shown in Fig. 5.30b. Like the GP zones in Fig. 5.30a, the θ'' precipitates are visible by virtue of the coherency-strain fields caused by the misfit perpendicular to the plates. θ'' precipitates are larger than GP zones being up to ~10 nm thick and 100 nm in diameter.

θ' is also tetragonal with an approximate composition CuAl_2 and again has (001) planes that are identical with $\{001\}_\alpha$. The (100) and (010) planes,

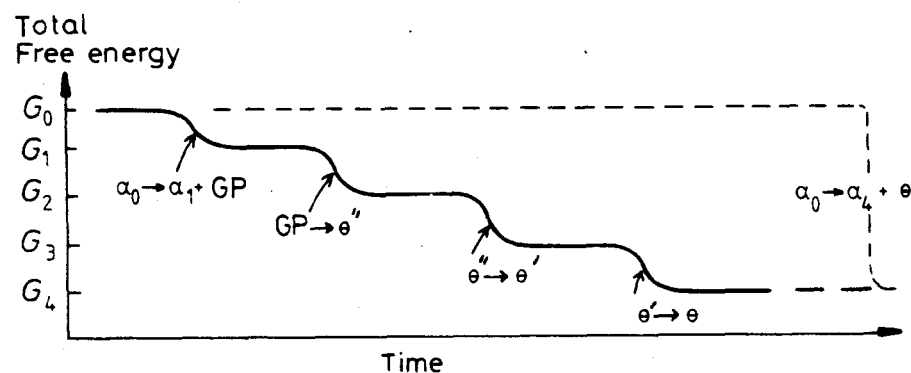
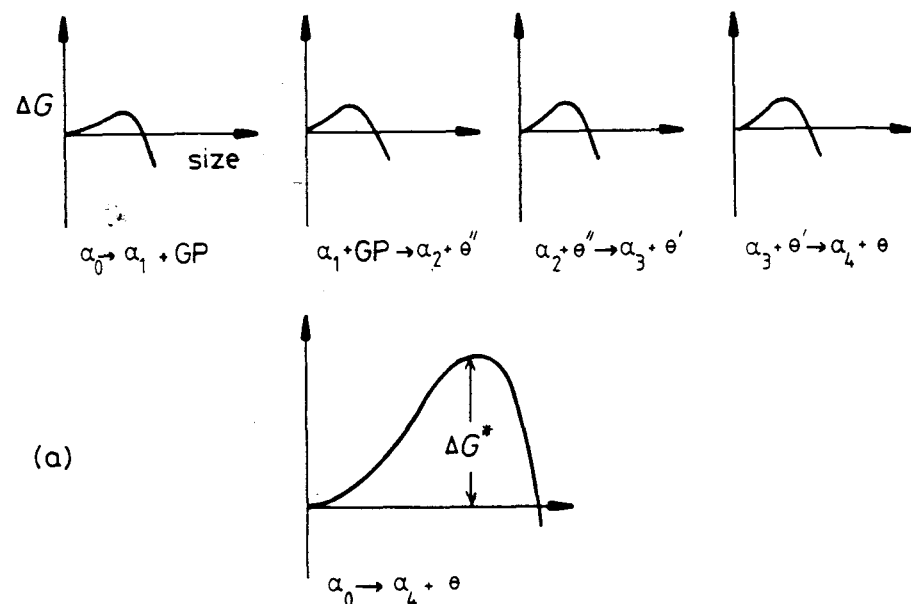


Fig. 5.28 (a) The activation energy barrier to the formation of each transition phase is very small in comparison to the barrier against the direct precipitation of the equilibrium phase. (b) Schematic diagram showing the total free energy of the alloy v. time.

however, have a different crystal structure to the matrix and a large misfit in the [001] direction. θ' therefore forms as plates on $\{001\}_\alpha$ with the same orientation relationship as θ'' . The broad faces of the plates are initially fully coherent but lose coherency as the plates grow, while the edges of the plates are either incoherent or have a complex semicoherent structure. A transmission electron micrograph of θ' plates $\sim 1 \mu\text{m}$ diameter is shown in Fig. 5.30c. Note the presence of misfit dislocations in the broad faces of the precipitates.

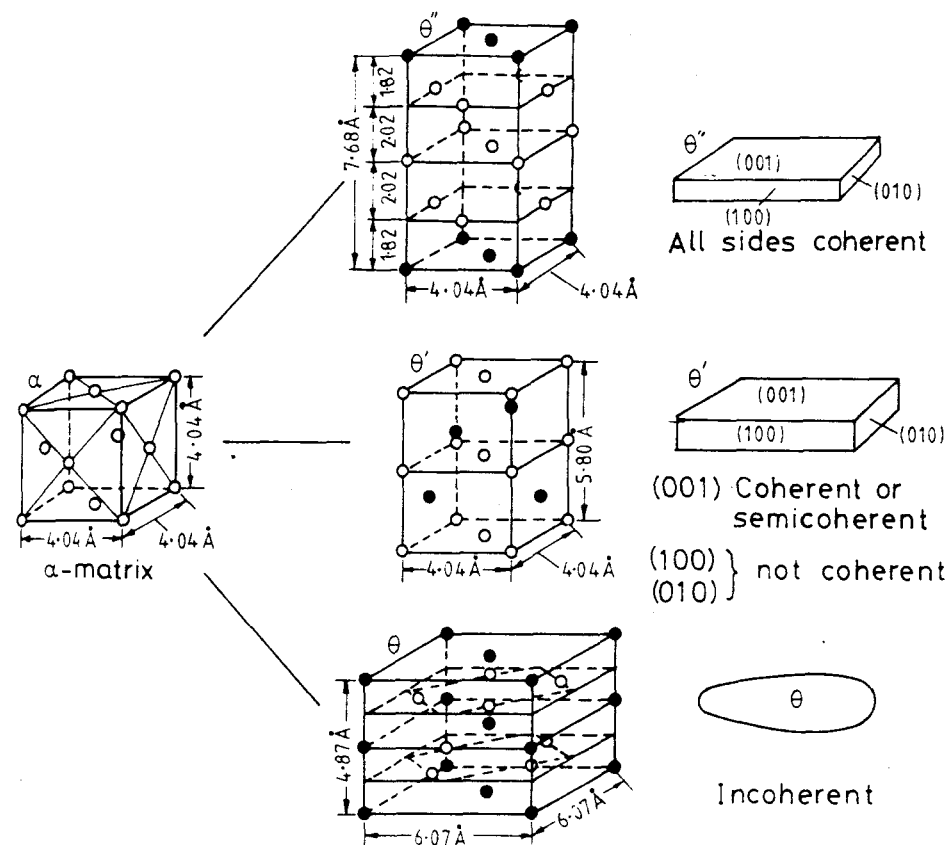


Fig. 5.29 Structure and morphology of θ'' , θ' and θ in Al-Cu ($\circ$ Al, $\bullet$ Cu).

Note also that since the edges of the plates are not coherent there are no long-range coherency-strain fields.

The equilibrium θ phase has the approximate composition CuAl_2 and a complex body-centred tetragonal structure as shown in Fig. 5.29. There are no planes of good matching with the matrix and only incoherent, or at best complex semicoherent interfaces are possible. The microstructure at this final stage of ageing is shown in Fig. 5.30d. Note the large size and coarse distribution of the precipitates.

The transformation from GP zones to θ'' occurs by the *in situ* transformation of the zones, which can be considered as very potent nucleation sites for θ'' . After longer ageing times the θ' phase nucleates on matrix dislocations with two orientations of θ' plates on any one $\frac{1}{2}\langle 110 \rangle$ dislocation. This is because the strain field of such a dislocation is able to reduce the misfit in two $\langle 100 \rangle$ matrix directions. Figure 5.31a shows θ' plates that have nucleated on dislocations. Note that as the θ' grows the surrounding, less-stable θ'' can be seen to dissolve. After still longer ageing times the equilibrium θ phase

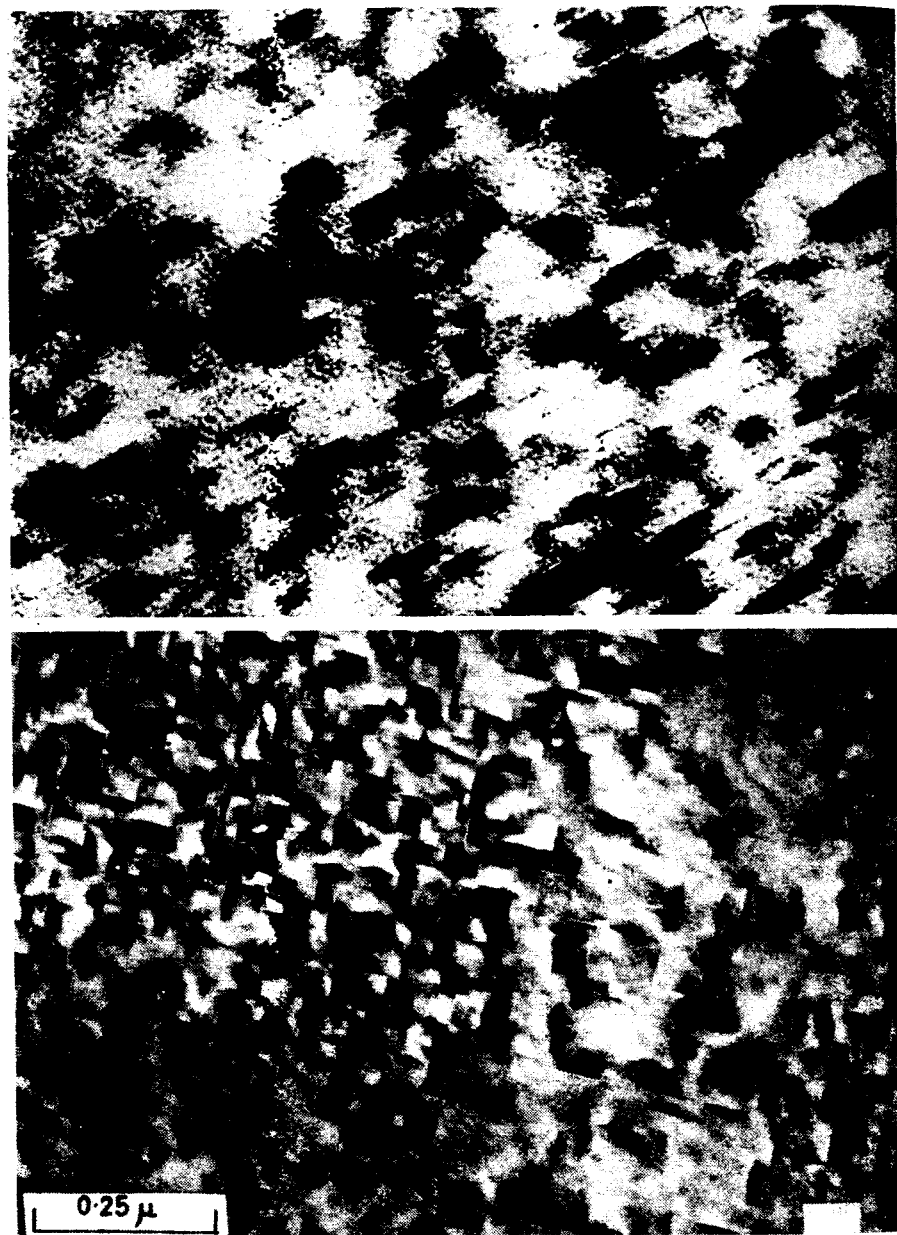


Fig. 5.30 Microstructures at different stages during ageing of Al-Cu alloys. (a) GP zones $\times 720\,000$. (b) $\theta'' \times 63\,000$. (c) $\theta' \times 18\,000$. (d) $\theta \times 8000$. [(a) After R.B. Nicholson and J. Nutting, *Philosophical Magazine* 3 (1958) 531. (b) R.B. Nicholson, G. Thomas and J. Nutting, *Journal of the Institute of Metals* 87 (1958-1959) 431. (c) G.C. Weatherly and R.B. Nicholson, *Philosophical Magazine* 17 (1968) 813. (d) G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972, from C. Laird.]



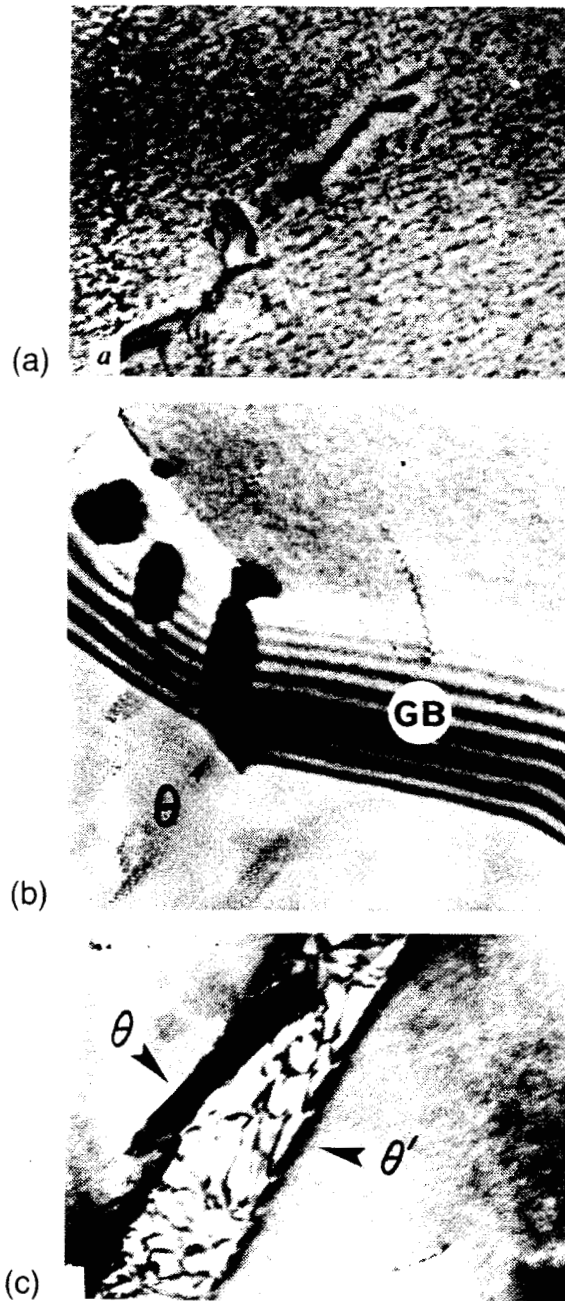


Fig. 5.31 Electron micrographs showing nucleation sites in Al-Cu alloys. (a) $\theta'' \rightarrow \theta'$. θ' nucleates at dislocation ($\times 70\,000$). (b) θ nucleation on grain boundary (GB) ($\times 56\,000$). (c) $\theta' \rightarrow \theta$. θ nucleates at $\theta'/$ matrix interface ($\times 70\,000$). (After P. Haasen, *Physical Metallurgy*, Cambridge University Press, Cambridge, 1978.)

nucleates either on grain boundaries, Fig. 5.31b, or at $\theta'/$ matrix interfaces, Fig. 5.31c. The choice of these nucleation sites is governed by the need to reduce the large interfacial energy contribution to ΔG^* for this phase.

The full sequence of GP zones and transition precipitates is only possible when the alloy is aged at a temperature below the GP zones solvus. For example, if ageing is carried out at a temperature above the θ'' solvus but below the θ' solvus, Fig. 5.25, the first precipitate will be θ' , heterogeneously nucleated on dislocations. If ageing is carried out above the θ' solvus, the only precipitate that is possible is θ which nucleates and grows at grain boundaries. Also, if an alloy containing GP zones is heated to above the GP zone solvus the zones will dissolve. This is known as *reversion*.

The effect of ageing temperature on the sequence of precipitates is illustrated by a schematic TTT diagram in Fig. 5.32. The fastest transformation rates are associated with the highest nucleation rates and therefore the finest precipitate distributions. There is consequently an increasing coarseness of microstructure through the sequence of precipitates as can be seen in Fig. 5.30.

The mechanism whereby a more stable precipitate grows at the expense of a less stable precipitate is illustrated in Fig. 5.33 for the case θ''/θ' . Figure 5.27 shows that the Cu concentration in the matrix close to the θ'' precipitates (α_2) will be higher than that close to θ' (α_3). Therefore Cu will tend to diffuse through the matrix away from θ'' , which thereby dissolves, and towards θ' , which grows.

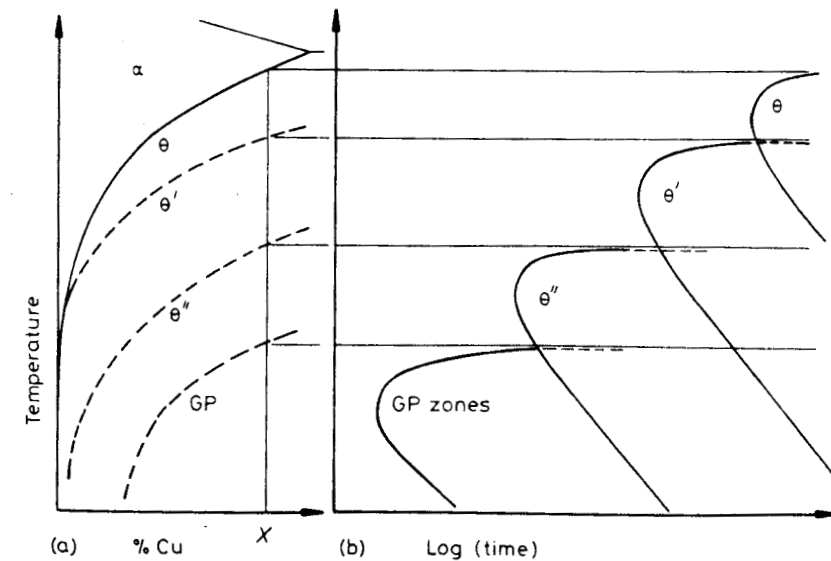


Fig. 5.32 (a) Metastable solvus lines in Al-Cu (schematic). (b) Time for start of precipitation at different temperatures for alloy X in (a).

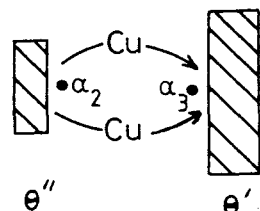


Fig. 5.33 Matrix in equilibrium with θ'' (α_2) contains more Cu than matrix in equilibrium with θ' (α_3). Cu diffuses as shown causing θ'' to shrink and θ' to grow.

5.5.2 Precipitation in Aluminium–Silver Alloys

Figure 5.34 shows the Al–Ag phase diagram. If alloys containing up to about 23 atomic % Ag are solution treated, quenched and given a low-temperature ageing treatment the precipitation sequence is

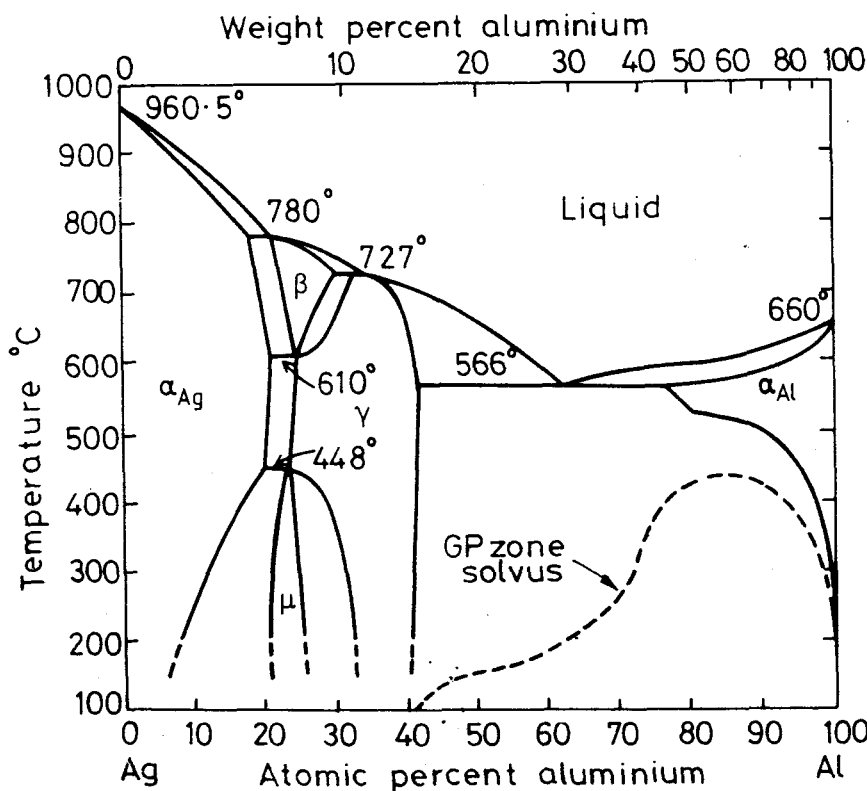
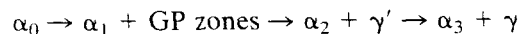


Fig. 5.34 Al–Ag phase diagram showing metastable two-phase field corresponding to GP zones. (After R. Baur and V. Gerold, *Zeitschrift für Metallkunde* 52 (1961) 671.)

As discussed earlier, the GP zones in this system are spherical. γ' is a close-packed hexagonal transition phase with an orientation relationship to the matrix of

$$(0001)_{\gamma'} // (111)_{\alpha}$$

$$[11\bar{2}0]_{\gamma'} // [1\bar{1}0]_{\alpha}$$

γ' is heterogeneously nucleated on helical dislocations by the enrichment of stacking faults with silver as discussed in Section 5.2. The equilibrium γ phase has the composition Ag_2Al , is hexagonal and has the same orientation relationship with the matrix as γ' . It forms as plate-like precipitates with (111) habit planes. γ can be formed from γ' by the latter acquiring misfit dislocations. It can also be separately nucleated at grain boundaries and grow by a cellular mechanism (see Section 5.7).

5.5.3 Quenched-in Vacancies

It was shown in Chapter 1 that the equilibrium concentration of vacancies increases exponentially with temperature. Thus the *equilibrium* vacancy concentration will be relatively high at the solution treatment temperature and much lower at the ageing temperature. However, when the alloy is rapidly quenched from the high temperature there will be no time for the new equilibrium concentration to be established and the high vacancy concentration becomes *quenched-in*. Given time, those vacancies in excess of the equilibrium concentration will *anneal out*. There will be a tendency for vacancies to be attracted together into *vacancy clusters*, and some clusters collapse into dislocation loops which can grow by absorbing more vacancies. The dislocations that are already present can also absorb vacancies by *climbing*. In this way straight screw dislocations can become converted into longer helical edge dislocations. There are many ways, therefore, in which excess vacancies are able to provide heterogeneous nucleation sites.

Another effect of quenched-in vacancies is to greatly increase the rate at which atoms can diffuse at the ageing temperatures. This in turn speeds up the process of nucleation and growth. Indeed the only way of explaining the rapid formation of GP zones at the relatively low ageing temperatures used is by the presence of excess vacancies.

If GP zones are separated by a mean spacing λ , the mean diffusion distance for the solute atoms is $\lambda/2$. Therefore, if the zones are observed to form in a time t , the effective diffusion coefficient is roughly given by x^2/t , i.e.

$$D \approx \frac{\lambda^2}{4t}$$

If high-temperature diffusion data are extrapolated down to the ageing temperature, the values obtained are orders of magnitude smaller than the above value. The difference can, however, be explained by a quenched-in

vacancy concentration that is orders of magnitude greater than the equilibrium value. In Al-Cu alloys, for example, GP zones can form by ageing at room temperature, which would not be feasible without assistance from excess vacancies.

There is other evidence for the role of quenched-in vacancies in enhancing diffusion rates. If the alloy is quenched from different solution treatment temperatures and aged at the same temperature, the initial rate of zone formation is highest in the specimens quenched from the highest temperatures. Also, if the quench is interrupted at an intermediate temperature, so that a new equilibrium concentration can be established, the rate of transformation is reduced. Reducing the rate of cooling from the solution treatment temperature produces a similar effect by allowing more time for vacancies to be lost during the quench. This is important when large parts are to be heat treated as the cooling rate varies greatly from the surface to the centre when the specimen is water-quenched for example.

Apart from dislocations, the main sinks for excess vacancies are the grain boundaries and other interfaces within the specimen. Since vacancies have such a high diffusivity it is difficult to avoid losing vacancies in the vicinity of grain boundaries and interfaces. This has important effects on the distribution of precipitates that form in the vicinity of grain boundaries on subsequent ageing. Figure 5.35a shows the vacancy concentration profiles that should be produced by vacancy diffusion to grain boundaries during quenching. Close to the boundary the vacancy concentration will be the equilibrium value for the ageing temperature, while away from the boundary it will be that for the solution treatment temperature. On ageing these alloys it is found that a precipitate-free zone (PFZ) is formed as shown in Fig. 5.35b. The solute concentration within the zone is largely unchanged, but no nucleation has occurred. The reason for this is that a *critical vacancy supersaturation* must be exceeded for nucleation to occur. The width of the PFZ is determined by the vacancy concentration as shown in Fig. 5.35c. At low temperatures, where the driving force for precipitation is high, the critical vacancy supersaturation is lower and narrower PFZs are formed. High quench rates will also produce narrow PFZs by reducing the width of the vacancy concentration profile. Similar PFZs can also form at inclusions and dislocations.

Finally, it should be mentioned that another cause of PFZs can be the nucleation and growth of grain boundary precipitates during cooling from the solution treatment temperature. This causes solute to be drained from the surrounding matrix and a PFZ results. An example of this type of PFZ is shown in Fig. 5.36.

5.5.4 Age Hardening

The reason for the interest in alloy systems that show transition phase precipitation is that great improvements in the mechanical properties of these alloys can be achieved by suitable solution treatment and ageing operations.

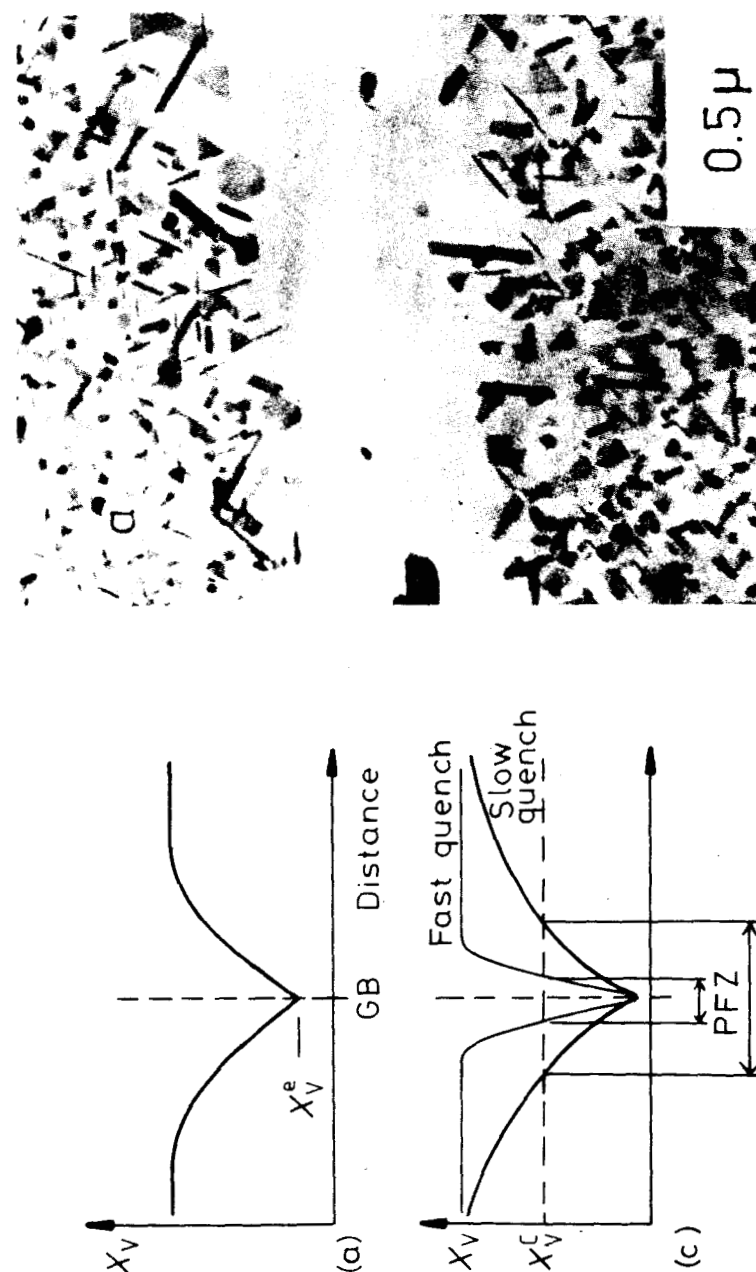


Fig. 5.35 A PFZ due to vacancy diffusion to a grain boundary during quenching. (a) Vacancy concentration profile. (b) A PFZ in an Al-Ge alloy ($\times 20\,000$). (c) Dependence of PFZ width on critical vacancy concentration X_v^c and rate of quenching. [(b) After G. Lorimer in *Precipitation in Solids*, K.C. Russell and H.I. Aaronson (Eds.), The Metallurgical Society of AIME, 1978.]

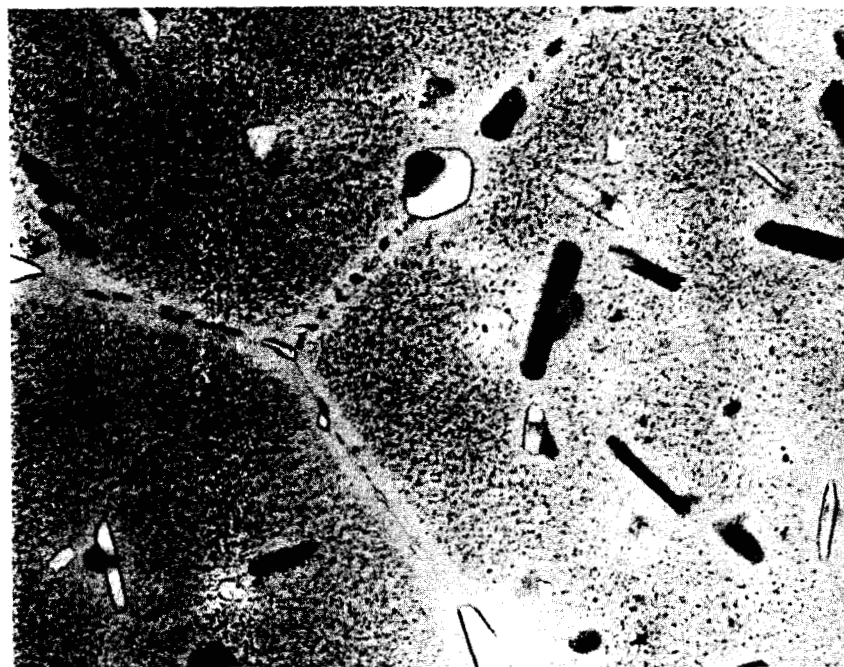


Fig. 5.36 PFZs around grain boundaries in a high-strength commercial Al-Zn-Mg-Cu alloy. Precipitates on grain boundaries have extracted solute from surrounding matrix. ($\times 59\,200$)

This is illustrated for various Al-Cu alloys in Fig. 5.37. The alloys were solution treated in the single-phase α region of the phase diagram, quenched to room temperature and aged at either 130 °C (Fig. 5.37a) or 190 °C (Fig. 5.37b). The curves show how the hardness of the specimens varies as a function of time and the range of time over which GP zones, θ'' and θ' appear in the microstructure. Immediately after quenching the main resistance to dislocation movement is solid solution hardening. The specimen is relatively easily deformed at this stage and the hardness is low. As GP zones form the hardness increases due to the extra stress required to force dislocations through the coherent zones.

The hardness continues to increase with the formation of the coherent θ'' precipitates because now the dislocations must also be forced through the highly strained matrix that results from the misfit perpendicular to the θ'' plates (see Fig. 5.30b). Eventually, with the formation of θ' the spacing between the precipitates becomes so large that the dislocations are able to bow between the precipitates and the hardness begins to decrease. Maximum hardness is associated with a combination of θ'' and θ' . Further ageing increases the distance between the precipitates making dislocation bowing easier and the hardness decreases. Specimens aged beyond peak hardness are referred to as *overaged*.

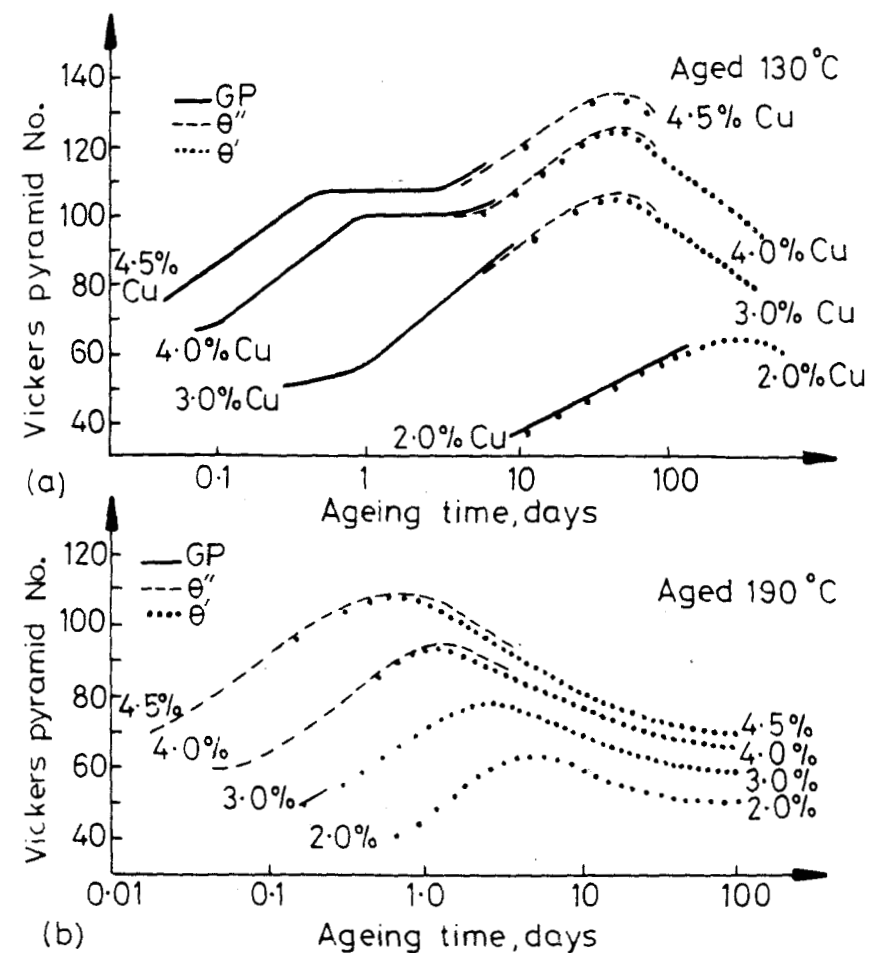


Fig. 5.37 Hardness v. time for various Al-Cu alloys at (a) 130 °C (b) 190 °C. (After J.M. Silcock, T.J. Heal and H.K. Hardy, *Journal of the Institute of Metals* 82 (1953-1954) 239.)

If Al-4.5 wt% Cu is aged at 190 °C, GP zones are unstable and the first precipitate to form is θ'' . The volume fraction of θ'' increases with time causing the hardness to increase as shown in Fig. 5.37b. However, at 190 °C the θ'' nucleates under the influence of a smaller driving force than at 130 °C and the resultant precipitate dispersion is therefore coarser. Also the maximum volume fraction of θ'' is reduced. Both of these factors contribute to a lower peak hardness on ageing at the higher temperature (compare Fig. 5.37a and b). However, diffusion rates are faster at higher temperatures and peak hardness is therefore achieved after shorter ageing times.

It can be seen that at 130 °C peak hardness in the Al-4.5 wt% Cu alloy is not reached for several tens of days. The temperatures that can be used in the

heat treatment of commercial alloys are limited by economic considerations to those which produce the desired properties within a reasonable period of time, usually up to ~ 24 h. In some high-strength alloys use is, therefore, made of a double ageing treatment whereby ageing is carried out in two steps: first at a relatively low temperature below the GP zone solvus, and then at a higher temperature. In this way a fine dispersion of GP zones obtained during the first stage can act as heterogeneous nucleation sites for precipitation at the higher temperature. This type of treatment can lead to a finer precipitate distribution than would be obtained from a single ageing treatment at the higher temperature.

Another treatment used commercially is to give the alloy a controlled deformation either before a single-stage age or between the two stages of a double-ageing treatment. The strength of the alloy after this treatment can be increased by a higher precipitate density, resulting from a higher nucleation rate, and by the retained dislocation networks which also act as a barrier to further deformation. However, deformation prior to ageing does not always result in an improvement in properties. In some cases deformation can lead to a coarser precipitate distribution.

Precipitation hardening is common to many alloy systems. Some of the more important systems are listed in Table 5.2. Some commercial alloys are listed in Table 5.3, along with their mechanical properties. In many of these systems it is possible to come very close to the maximum theoretical strength of the matrix, i.e. about $\sim \mu/30$. However, engineering alloys are not heat treated for maximum strength alone. Consideration must also be given to toughness, stress corrosion resistance, fatigue, etc., when deciding on the best heat treatment in practice.

5.5.5 Spinodal Decomposition

It was mentioned at the beginning of this chapter that there are certain transformations where there is no barrier to nucleation. One of these is the spinodal mode of transformation. Consider a phase diagram with a miscibility gap as shown in Fig. 5.38a. If an alloy with composition X_0 is solution treated at a high temperature T_1 and then quenched to a lower temperature T_2 the composition will initially be the same everywhere and its free energy will be G_0 on the G curve in Fig. 5.38b. However, the alloy will be immediately *unstable* because small fluctuations in composition that produce A-rich and B-rich regions will cause the total free energy to decrease. Therefore 'up-hill' diffusion takes place as shown in Fig. 5.39 until the equilibrium compositions X_1 and X_2 are reached.

The above process can occur for any alloy composition where the free energy curve has a negative curvature, i.e.

$$\frac{d^2G}{dX^2} < 0 \quad (5.42)$$

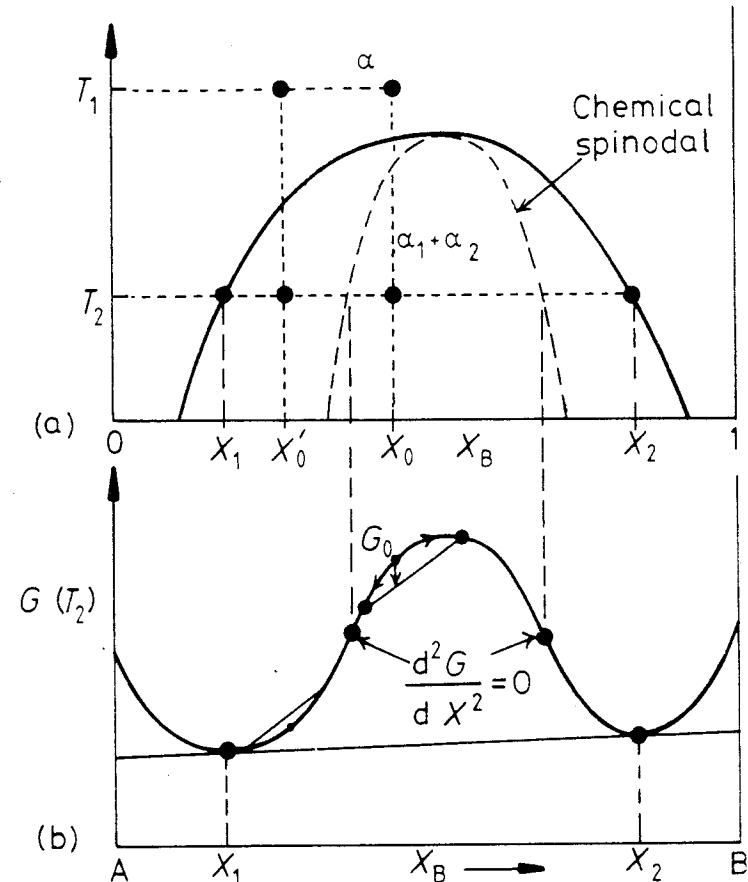


Fig. 5.38 Alloys between the spinodal points are unstable and can decompose into two coherent phases α_1 and α_2 without overcoming an activation energy barrier. Alloys between the coherent miscibility gaps and the spinodal are metastable and can decompose only after nucleation of the other phase.

Therefore the alloy must lie between the two points of inflection on the free energy curve. The locus of the points on the phase diagram, Fig. 5.32a, is known as the chemical spinodal.

If the alloy lies outside the spinodal, small variations in composition lead to an increase in free energy and the alloy is therefore *metastable*. The free energy of the system can only be decreased in this case if nuclei are formed with a composition very different from the matrix. Therefore, outside the spinodal the transformation must proceed by a process of nucleation and growth. Normal down-hill diffusion occurs in this case as shown in Fig. 5.40.

The rate of spinodal transformation is controlled by the interdiffusion coefficient, D . Within the spinodal $D < 0$ and the composition fluctuations shown in Fig. 5.39 will therefore increase exponentially with time, with a

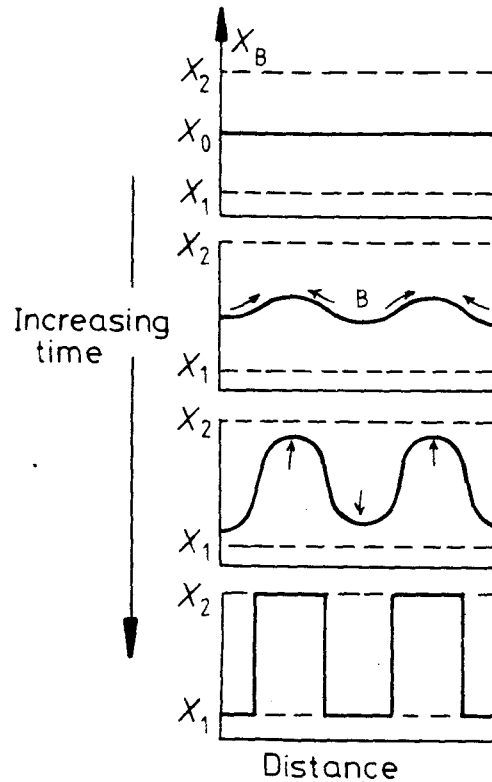


Fig. 5.39 Schematic composition profiles at increasing times in an alloy quenched into the spinodal region (X_0 in Fig. 5.38).

characteristic time constant $\tau = -\lambda^2/4\pi^2D$, where λ is the wavelength of the composition modulations (assumed one-dimensional). The rate of transformation can therefore become very high by making λ as small as possible. However, as will be shown below, there is a minimum value of λ below which spinodal decomposition cannot occur.

In order to be able to calculate the wavelength of the composition fluctuations that develop in practice it is necessary to consider two important factors that have been omitted from the above discussion: (1) interfacial energy effects, and (2) coherency strain energy effects.

If a homogeneous alloy of composition X_0 decomposes into two parts one with composition $X_0 + \Delta X$ and the other with composition $X_0 - \Delta X$, it can be shown that⁷ the total chemical free energy will change by an amount ΔG_c given by

$$\Delta G_c = \frac{1}{2} \frac{d^2G}{dX^2} (\Delta X)^2 \quad (5.43)$$

If, however, the two regions are finely dispersed and coherent with each

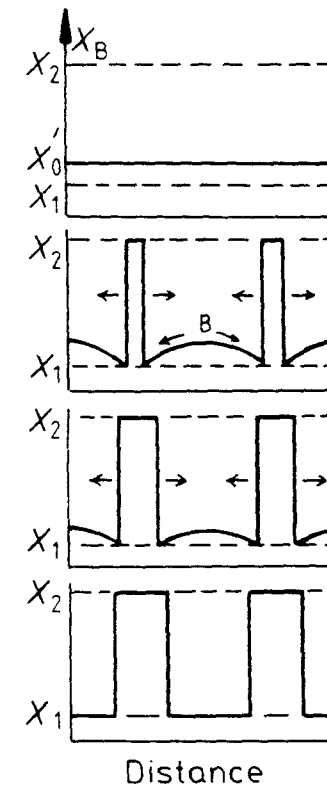


Fig. 5.40 Schematic composition profiles at increasing times in an alloy outside the spinodal points (X_0 in Fig. 5.38).

other there will be an additional energy change due to interfacial energy effects. Although, during the early stages of spinodal decomposition, the interface between A-rich and B-rich regions is not sharp but very diffuse, there is still an effective interfacial energy contribution. The magnitude of this energy depends on the composition gradient across the interface, and for this reason it is known as a 'gradient energy'. In solid solutions which tend to cluster the energy of like atom-pairs is less than that of unlike pairs. Thus the origin of the gradient energy is the increased number of unlike nearest neighbours in a solution containing composition gradients compared to a homogeneous solution. For a sinusoidal composition modulation of wavelength λ and amplitude ΔX the maximum composition gradient is proportional to $(\Delta X/\lambda)$ and the gradient energy term ΔG_γ is given by

$$\Delta G_\gamma = K \left(\frac{\Delta X}{\lambda} \right)^2 \quad (5.44)$$

where K is a proportionality constant dependent on the difference in the bond energies of like and unlike atom pairs.

If the sizes of the atoms making up the solid solution are different, the generation of composition differences will introduce a coherency strain energy term, ΔG_s . If the misfit between the A-rich and B-rich regions is δ , $\Delta G_s \propto E\delta^2$ where E is Young's modulus. For a total composition difference ΔX , δ will be given by $(da/dX)\Delta X/a$, where a is the lattice parameter. An exact treatment of the elastic strain energy shows that

$$\Delta G_s = \eta^2(\Delta X)^2 E' V_m \quad (5.45)$$

where

$$\eta = \frac{1}{a} \left(\frac{da}{dX} \right) \quad (5.46)$$

i.e. η is the fractional change in lattice parameter per unit composition change. $E' = E/(1 - \nu)$, where ν is Poisson's ratio, and V_m is the molar volume. Note that ΔG_s is independent of λ .

If all of the above contributions to the total free energy change accompanying the formation of a composition fluctuation are summed we have

$$\Delta G = \left\{ \frac{d^2 G}{dX^2} + \frac{2K}{\lambda^2} + 2\eta^2 E' V_m \right\} \frac{(\Delta X)^2}{2} \quad (5.47)$$

It can be seen therefore that the condition for a homogeneous solid solution to be unstable and decompose spinodally is that

$$-\frac{d^2 G}{dX^2} > \frac{2K}{\lambda^2} + 2\eta^2 E' V_m \quad (5.48)$$

Thus the limits of temperature and composition within which spinodal decomposition is possible are given by the conditions $\lambda = \infty$ and

$$\frac{d^2 G}{dX^2} = -2\eta^2 E' V_m \quad (5.49)$$

The line in the phase diagram defined by this condition is known as the *coherent spinodal* and it lies entirely within the chemical spinodal ($d^2 G/dX^2 = 0$) as shown in Fig. 5.41. It can be seen from Equation 5.48 that the wavelength of the composition modulations that can develop inside the coherent spinodal must satisfy the condition

$$\lambda^2 > -2K / \left(\frac{d^2 G}{dX^2} + 2\eta^2 E' V_m \right) \quad (5.50)$$

Thus the minimum possible wavelength decreases with increasing undercooling below the coherent spinodal.

Figure 5.41 also shows the coherent miscibility gap. This is the line defining the equilibrium compositions of the coherent phases that result from spinodal decomposition (X_1 and X_2 in Fig. 5.39). The miscibility gap that normally

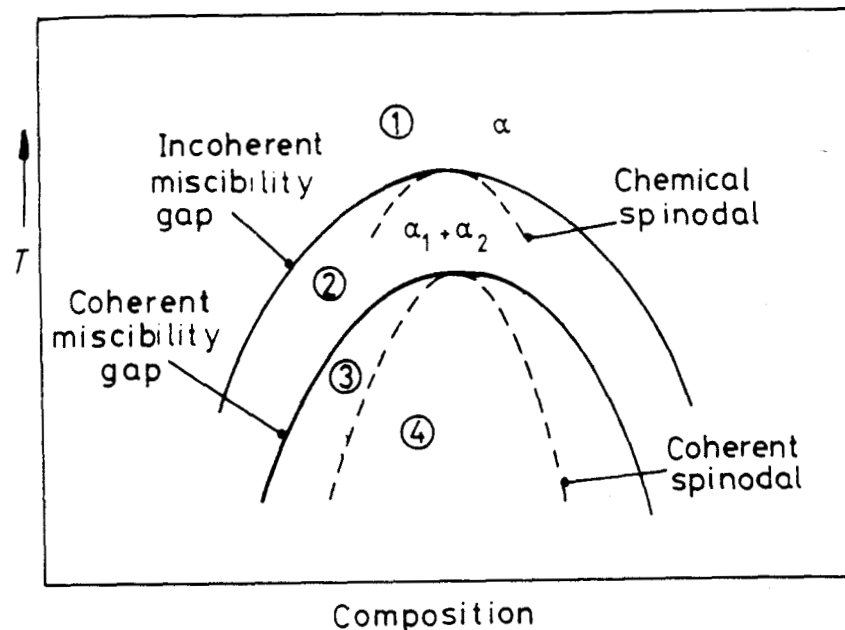


Fig. 5.41 Schematic phase diagram for a clustering system. Region 1: homogeneous α stable. Region 2: homogeneous α metastable, only incoherent phases can nucleate. Region 3: homogeneous α metastable, coherent phases can nucleate. Region 4: homogeneous α unstable, no nucleation barrier, spinodal decomposition occurs.

appears on an equilibrium phase diagram is the incoherent (or equilibrium) miscibility gap. This corresponds to the equilibrium compositions of *incoherent* phases, i.e. in the absence of strain fields. The chemical spinodal is also shown in Fig. 5.41 for comparison, but it is of no practical importance.

Spinodal decomposition is not only limited to systems containing a stable miscibility gap. All systems in which GP zones form, for example, contain a *metastable* coherent miscibility gap, i.e. the GP zone solvus (see the Al-Ag system in Fig. 5.34 for example). Thus it is possible that at high supersaturations GP zones are able to form by the spinodal mechanism. If ageing is carried out below the coherent solvus but outside the spinodal, GP zones can only form by a process of nucleation and growth, Fig. 5.40. Between the incoherent and coherent miscibility gap, Fig. 5.41, $\Delta G_v - \Delta G_s < 0$ and only incoherent strain-free nuclei can form.

The difference in temperature between the coherent and incoherent miscibility gaps, or the chemical and coherent spinodals in Fig. 5.41, is dependent on the magnitude of $|\eta|$. When there is a large atomic size difference $|\eta|$ is large and a large undercooling is required to overcome the strain energy effects. As discussed earlier large values of $|\eta|$ in cubic metals can be mitigated if the misfit strains are accommodated in the elastically soft $\langle 100 \rangle$ directions. This is achieved by the composition modulations building up parallel to $\{100\}$.

Figure 5.42 shows a spinodal structure in a specimen of Al-22.5 Zn-0.1 Mg (atomic %) solution treated at 400 °C and aged 20 h at 100 °C. The wavelength in the structure is 25 nm, but this is greater than the initial microstructure due to coarsening which occurs on holding long times at high temperatures.

5.5.6 Particle Coarsening⁸

The microstructure of a two-phase alloy is always unstable if the total interfacial free energy is not a minimum. Therefore a high density of small precipitates will tend to *coarsen* into a lower density of larger particles with a smaller total interfacial area. However, such coarsening often produces an undesirable degradation of properties such as a loss of strength or the disappearance of grain-boundary pinning effects (see Section 3.3.5). As with grain growth, the rate of coarsening increases with temperature and is of particular concern in the design of materials for high temperature applications.

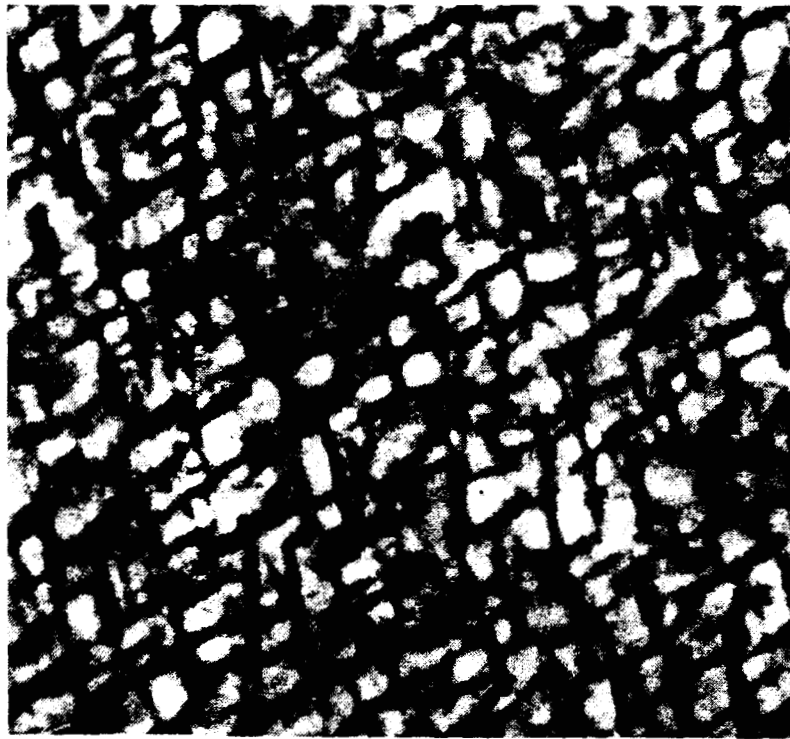


Fig. 5.42 A coarsened spinodal microstructure in Al-22.5 at% Zn-0.1 at% Mg solution treated 2 h at 400 °C and aged 20 h at 100 °C. Thin foil electron micrograph ($\times 314\,000$). (After K.B. Rundman, *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 184.)

In any precipitation-hardened specimen there will be a range of particle sizes due to differences in the time of nucleation and rate of growth. Consider two adjacent spherical precipitates⁹ with different diameters as shown in Fig. 5.43. Due to the Gibbs-Thomson effect, the solute concentration in the matrix adjacent to a particle will increase as the radius of curvature decreases. Fig. 5.43b. Therefore there will be concentration gradients in the matrix which will cause solute to diffuse in the direction of the largest particles away from the smallest, so that the small particles shrink and disappear while large particles grow. The overall result is that the total number of particles decreases and the mean radius ($\bar{r}$) increases with time. By assuming volume diffusion is the rate controlling factor it has been shown¹⁰ that the following relationship should be obeyed:

$$(\bar{r})^3 - r_0^3 = kt \quad (5.51)$$

where

$$k \propto D\gamma X_e$$

r_0 is the mean radius at time $t = 0$, D is the diffusion coefficient, γ is the interfacial energy and X_e is the equilibrium solubility of very large particles. Since D and X_e increase exponentially with temperature, the rate of coarsening will increase rapidly with increasing temperature. Fig. 5.44. Note that the rate of coarsening

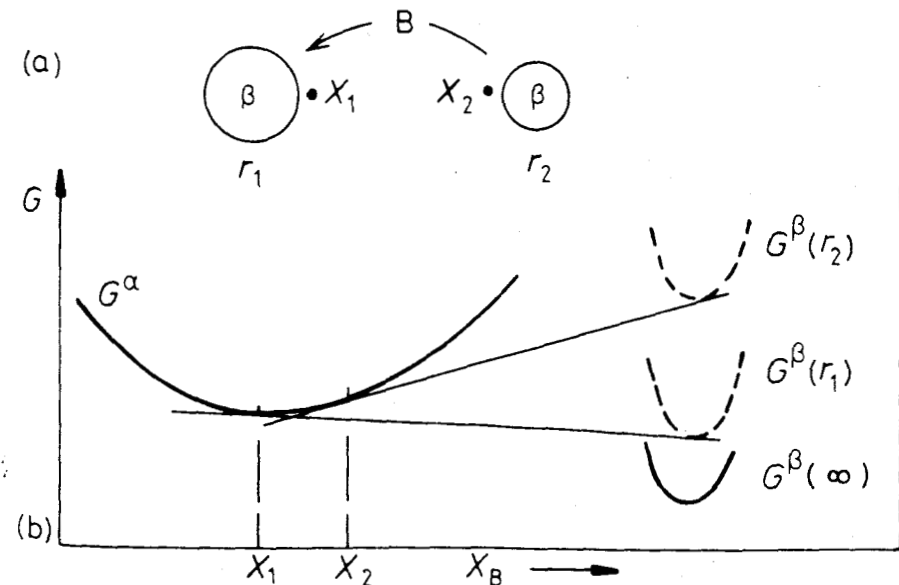


Fig. 5.43 The origin of particle coarsening. β with a small radius of curvature (r_2) has a higher molar free energy than β with a large radius of curvature (r_1). The concentration of solute is therefore highest outside the smallest particles.

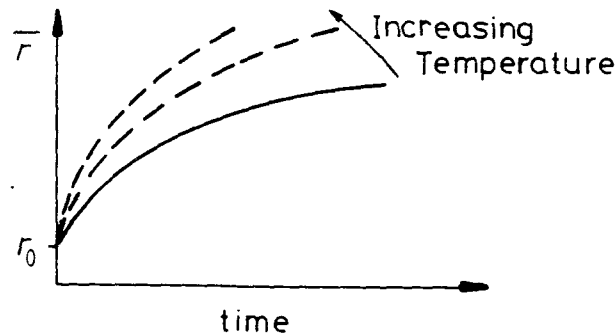


Fig. 5.44 Schematic diagram illustrating how the mean particle radius $\bar{r}$ increases with time at different temperatures.

$$\frac{d\bar{r}}{dt} \propto \frac{k}{\bar{r}^2} \quad (5.52)$$

so that distributions of small precipitates coarsen most rapidly.

In practice the rate at which particles coarsen may not follow a linear $\bar{r}^3-t$ relationship. Deviations from this relationship can be caused by diffusion short-circuits such as dislocations, or grain boundaries. Also the coarsening rate may be interface controlled. Nevertheless, apart from the case of interface control, the rate of coarsening should depend on the product $D\gamma X_e$, (k in Equation 5.51). Therefore high temperature alloys whose strength depends on a fine precipitate dispersion must have a low value for at least one of γ , X_e or D . Let us consider examples of each of these.

Low γ

The heat-resistant *Nimonic* alloys based on Ni–Cr with additions of Al and Ti obtain their high strength from a fine dispersion of the ordered fcc phase $\text{Ni}_3(\text{TiAl})$ (γ') which precipitates in the fcc Ni-rich matrix. The Ni/ γ' interfaces are fully coherent and the interfacial energy is exceptionally low (~ 10 – 30 mJ m^{-2}) which enables the alloys to maintain a fine structure at high temperature. The misfit between the precipitates and matrix varies between zero and about 0.2% depending on composition. It is interesting that the total creep–rupture life of these alloys can be increased by a factor of 50× by careful control of composition to give zero misfit as compared to 0.2% misfit. The reason for this may be that during creep deformation the particles with the slightly higher misfits lose coherency with the result that γ is increased thereby increasing the rate of coarsening.

Low X_e

High strength at high temperatures can also be obtained with fine oxide dispersions in a metal matrix. For example W and Ni can be strengthened for

high temperature use by fine dispersions of thorium ThO_2 . In general, oxides are very insoluble in metals and the stability of these microstructures at high temperatures can be attributed to a low value of X_e in the product $D\gamma X_e$.

Low D

Cementite dispersions in tempered steels coarsen very quickly due to the high diffusivity of interstitial carbon. However, if the steel contains a substitutional alloying element that segregates to the carbide, the rate of coarsening becomes limited by the much slower rate at which substitutional diffusion can occur. If the carbide-forming element is present in high concentrations more stable carbides are formed which have the additional advantage of a lower solubility (X_e). Therefore low-alloy steels used for medium temperature creep resistance often have additions of strong carbide-forming elements.

5.6 The Precipitation of Ferrite from Austenite

In this section we will be concerned with phase transformations in which the first phase to appear is that given by the equilibrium phase diagram. The discussion will be illustrated by reference to the diffusional transformation of Fe–C–austenite into ferrite. However, many of the principles are quite general and have analogues in other systems where the equilibrium phases are not preceded by the precipitation of transition phases. Under these conditions the most important nucleation sites are grain boundaries and the surfaces of inclusions.

Consider an Fe–0.15 wt% C alloy which, after austenitizing, is allowed to partially transform to ferrite at various temperatures below A_3 (Fig. 5.45) and then quenched into water. The resultant microstructures are shown in Fig. 5.46. The white areas are ferrite (α). The grey areas are martensite that formed from the untransformed austenite (γ) during quenching. At small undercooling below A_3 , Fig. 5.46a, the ferrite nucleates on austenite grain boundaries and grows in a 'blocky' manner to form what are known as grain-boundary allotriomorphs. Note that both smoothly curved, presumably incoherent, α/γ interfaces as well as faceted, semicoherent interfaces are present. At larger undercoolings there is an increasing tendency for the ferrite to grow from the grain boundaries as plates, so-called Widmanstätten side-plates, which become finer with increasing undercooling, Fig. 5.46b, c and d.

Experimental measurements on Widmanstätten ferrite in other ferrous alloys show that the habit planes are irrational, scattered 4 to 20° from $\{111\}_\gamma$, and that orientation relationships close to the Nishiyama–Wasserman or Kurdjumov–Sachs type are usually found. High resolution transmission electron microscopy has also shown that the habit planes have a complex semicoherent structure, containing structural ledges and misfit dislocations, similar to that described in Section 3.4.1¹¹.

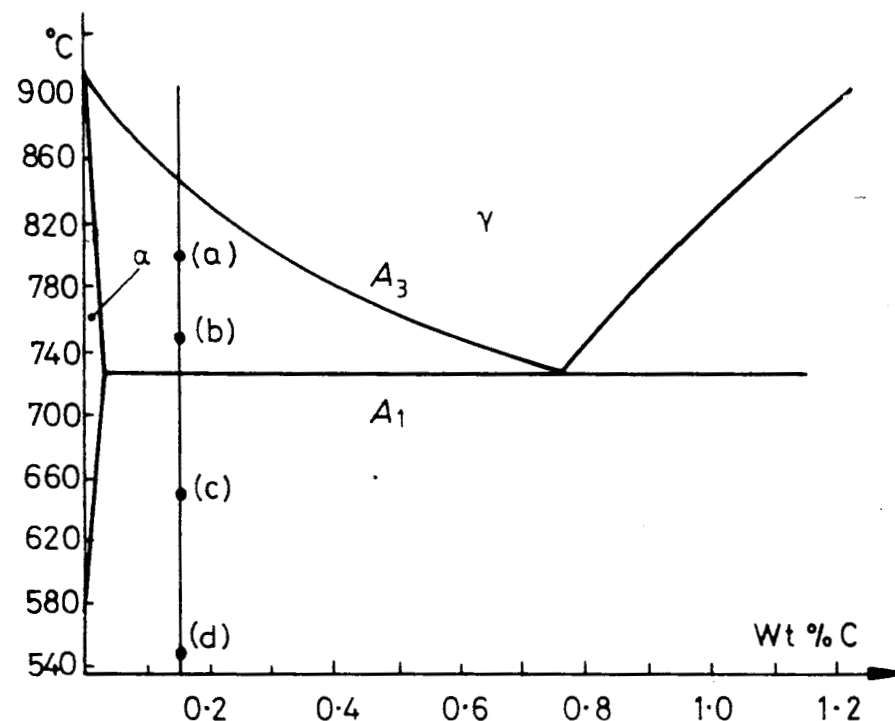


Fig. 5.45 Holding temperatures for steel in Fig. 5.46.

As explained previously, the need to minimize ΔG^* leads to the creation of semicoherent interfaces and orientation relationships, even in the case of grain-boundary nucleation. A critical nucleus could therefore appear as shown in Fig. 3.45b with faceted (planar) coherent (or semicoherent) interfaces and smoothly curved incoherent interfaces. For certain misorientations across the grain boundary it may even be possible for low-energy facets to form with both grains. Due to their low mobility faceted interfaces will tend to persist during growth while incoherent interfaces will be able to grow continuously and thereby retain a smooth curvature. Thus it is possible to explain the presence of smoothly curved and faceted interfaces in Fig. 5.46a.

The reason for the transition from grain boundary allotriomorphs to Widmanstätten side-plates with increasing undercooling is not fully understood. It has been suggested by Aaronson and co-workers¹² that the relative rates at which semicoherent and incoherent interfaces can migrate vary with undercooling as shown in Fig. 5.47. At small undercoolings it is proposed that both semicoherent and incoherent interfaces can migrate at similar rates, while at large undercoolings only incoherent interfaces can make full use of the increased driving force. Consideration of Fig. 5.13 thus shows that approximately equiaxed morphologies should develop at low undercoolings while plate-like morphologies, with ever-increasing aspect ratios, should develop at

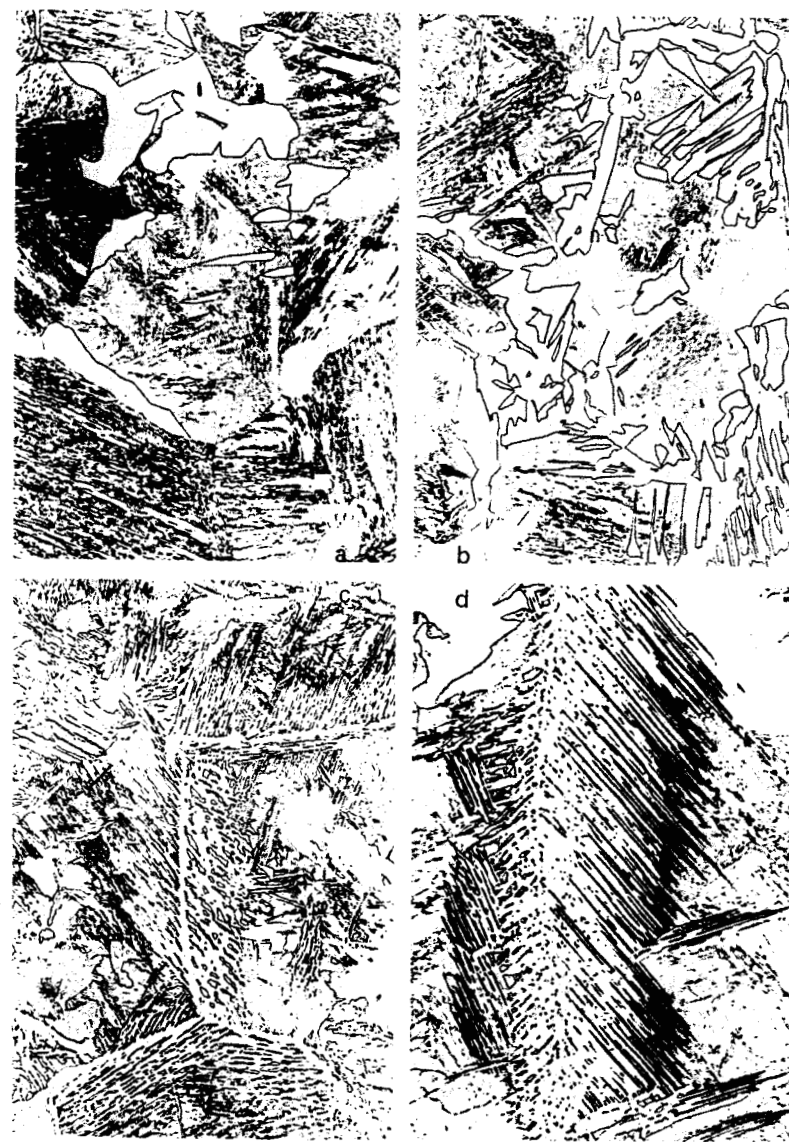


Fig. 5.46 Microstructures of an Fe-0.15% C alloy. The specimens were austenitized, held at an intermediate temperature to give some ferrite, and then quenched to room temperature. The ferrite is white. The grey, fine constituent is a mixture of ferrite and carbide formed on quenching. All photographs are $\times 100$ except (d). (a) 800 °C for 150 s—primarily ferrite allotriomorphs with a few plates. (b) 750 °C for 40 s—many more plates, mostly growing from grain boundaries. (c) 650 °C for 9 s—relatively fine. Note common direction of plates along each boundary. (d) 550 °C for 2 s ($\times 300$) (After P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, New York, 1969, after H.I. Aaronson.)

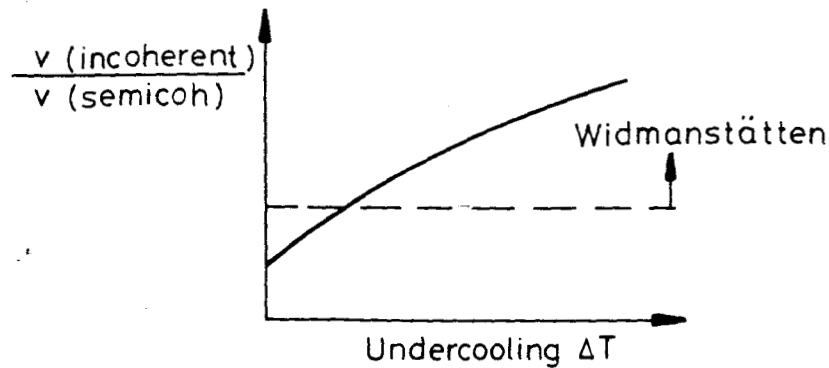


Fig. 5.47 A possible variation of the relative velocity of incoherent and semicoherent interfaces at different undercoolings. Above a certain ratio Widmanstätten morphologies should develop, as shown in Fig. 5.13. (After H.I. Aaronson, in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), 1962, by permission of The Metallurgical Society of AIME.)

high undercoolings. Another factor which may contribute to the increased fineness of the Widmanstätten morphologies with decreasing temperature is that the minimum plate-tip radius r^* is inversely proportional to the undercooling.

It can be seen in Fig. 5.46 that ferrite can also precipitate within the austenite grains (intragranular ferrite). Suitable heterogeneous nucleation sites are thought to be inclusions and dislocations. These precipitates are generally equiaxed at low undercoolings and more plate-like at higher undercoolings.

In general the nucleation rate within grains will be less than on grain boundaries. Therefore, whether or not intragranular precipitates are observed depends on the grain size of the specimen. In fine-grained austenite, for example, the ferrite that forms on grain boundaries will rapidly raise the carbon concentration within the middle of the grains, thereby reducing the undercooling and making nucleation even more difficult. In a large-grained specimen, however, it takes a longer time for the carbon rejected from the ferrite to reach the centres of the grains and meanwhile there will be time for nucleation to occur on the less favourable intragranular sites.

A TTT diagram for the precipitation of ferrite in a hypoeutectoid steel will have a typical C shape as shown in Fig. 5.48. The $\gamma \rightarrow \alpha$ transformation should be approximately described by Equation 5.39 and the time for a given percentage transformation will decrease as the constant k increases, e.g. Equation 5.40. As usual, k increases with small increases in T due to increased nucleation and growth rates— k is also raised by an increase in the total number of nucleation sites. Thus decreasing the austenite grain size has the effect of shifting the C curve to shorter transformation times.

It is possible to mark a temperature T_w below which the ferrite forms as predominantly Widmanstätten plates and above which it is mainly in the form

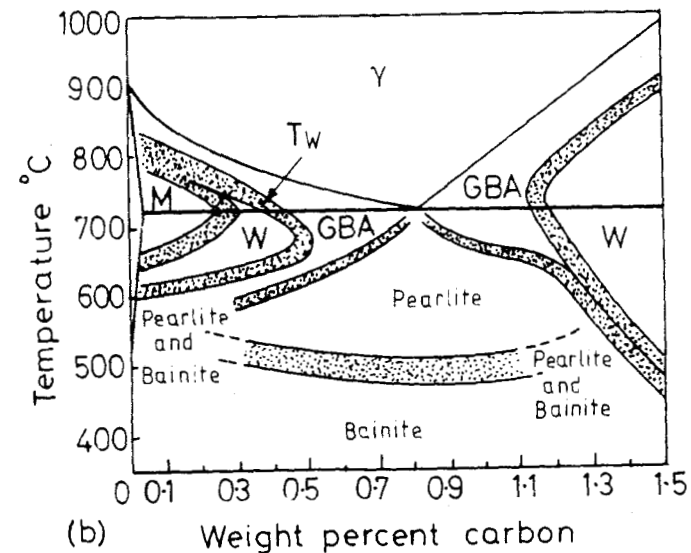
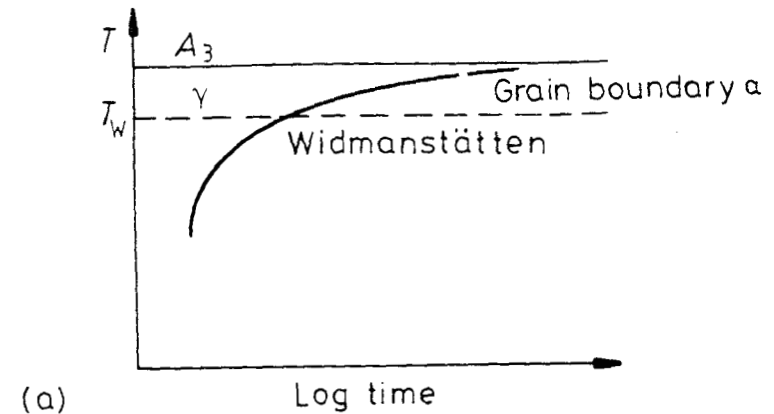


Fig. 5.48 (a) Typical TTT curve for $\gamma \rightarrow \alpha$ transformation. (b) Temperature-composition regions in which the various morphologies are dominant at late reaction times in specimens with ASTM grain size Nos. 0–1. GBA = grain boundary allotriomorphs, W = Widmanstätten sideplates and/or intragranular plates, M = massive ferrite, see Section 5.9. (After H.I. Aaronson, in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), 1962, by permission of The Metallurgical Society of AIME.)

of grain boundary allotriomorphs. For alloys of different carbon content A_3 and T_w vary as shown on the phase diagram in Fig. 5.48b.

During practical heat treatments, such as normalizing or annealing, transformation occurs continuously during cooling. Under these circumstances the final microstructure will depend on the cooling rate. If the specimen is cooled

very slowly there will be time for nucleation to occur at small undercoolings on grain corners, edges and boundaries. As these nuclei grow the carbon rejected into the austenite will have time to diffuse over large distances and the austenite grain should maintain a uniform composition given by the equilibrium phase diagram. Finally the austenite reaches the eutectoid composition and transforms to pearlite. Furnace cooling corresponds fairly closely to these conditions and an example is shown in Fig. 5.49c. The final proportions of ferrite and pearlite should be as determined by the equilibrium phase diagram.

The microstructure that results from more rapid cooling will depend on the grain size and the cooling rate. If the rate of cooling is moderately high the specimen will not remain long enough at high temperatures for nucleation to occur. Thus nuclei will not be formed until higher supersaturations are reached. The nucleation rate will then be rapid and large areas of grain boundary will become covered with nuclei. If the temperature is below T_w the ferrite will grow into the austenite as Widmanstätten side-plates with a spacing that becomes finer with decreasing temperature.

The nuclei that form at the highest temperatures will be on grain corners which will be followed by edges at lower temperatures and finally grain boundaries at still lower temperatures. In a *small-grained specimen* where there are a large number of grain corner and edge sites a large number of nuclei can be formed above the T_w temperature and grow as grain-boundary allotriomorphs. In a *larged-grained specimen*, on the other hand, relatively few nuclei will form at high temperatures and the austenite far from these particles will remain supersaturated until lower temperatures, below T_w , when ferrite will be able to nucleate on grain boundary sites and grow as Widmanstätten side-plates. The effect of cooling rate and grain size is illustrated in Fig. 5.49. Note also that the total volume fraction of ferrite decreases as the transformation temperature decreases. This point will be returned to later.

If the austenite contains more than about 0.8wt% C, the first phase to form will be cementite. This also nucleates and grows with an orientation relationship to the austenite, producing similar morphologies to ferrite—grain boundary allotriomorphs at high temperatures and Widmanstätten side-plates at lower temperatures as shown in Fig. 5.48b.

5.7 Cellular Precipitation

Grain-boundary precipitation does not always result in grain-boundary allotriomorphs or Widmanstätten side-plates or needles. In some cases it can result in a different mode of transformation, known as *cellular precipitation*. The essential feature of this type of transformation is that the boundary moves with the growing tips of the precipitates as shown in Fig. 5.50. Morphologically the transformation is very similar to the eutectoid reaction.

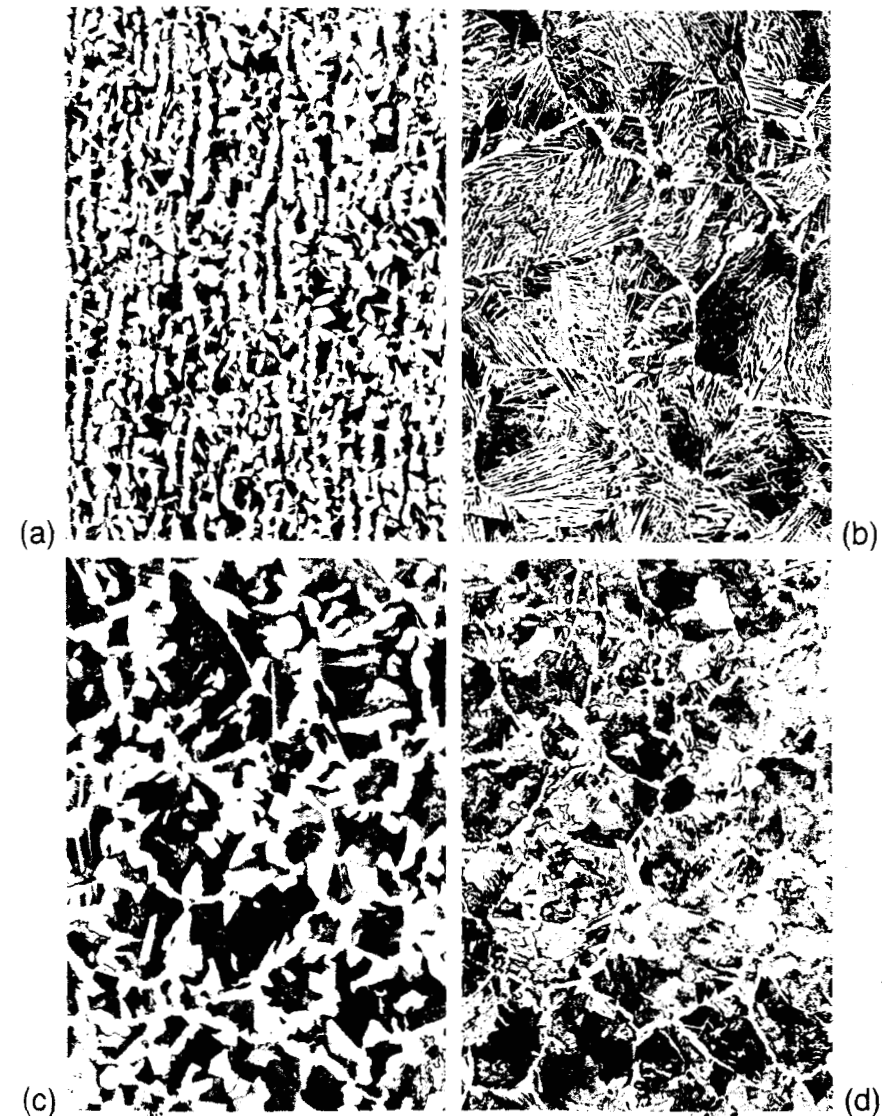


Fig. 5.49 Microstructures obtained from different heat treatments in plain carbon steels ($\times 60$), 0.23wt% C 1.2% Mn air-cooled, showing influence of prior austenite grain size: (a) austenitized at 900 °C (b) austenitized at 1150 °C. 0.4% C showing effect of cooling rate for same grain size: (c) furnace cooled (annealed), (d) air cooled (normalized). (After P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, New York, 1969: (a) and (b) after R. Yoe, (c) and (d) after K. Zurlippe.)

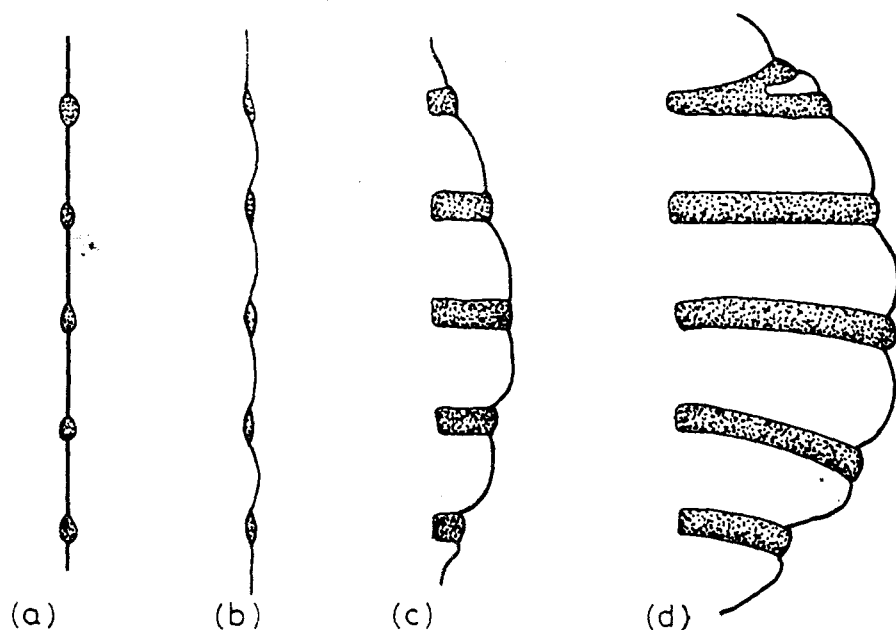
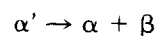


Fig. 5.50 A schematic diagram showing a possible sequence of steps during the development of cellular precipitation.

However, in this case the reaction can be written



where α' is the supersaturated matrix, α is the *same* phase but with a lower thermodynamic excess of solute, and β is the equilibrium precipitate. The mechanism whereby grain-boundary nucleation develops into cellular precipitation differs from one alloy to another and is not always fully understood. The reason why cells develop in some alloys and not in others is also unclear.

Figure 5.51 shows an example of cellular precipitation in a Mg–9 atomic % Al alloy. The β phase in this case is the equilibrium precipitate $\text{Mg}_{17}\text{Al}_{12}$ indicated in the phase diagram, Fig. 5.52. It can be seen in Fig. 5.51 that the $\text{Mg}_{17}\text{Al}_{12}$ forms as lamellae embedded in a Mg-rich matrix. The grain boundary between grains I and II was originally straight along AA but has been displaced, and the cell matrix and grain I are the same grain.

Figure 5.53 shows another specimen which has been given a two-stage heat treatment. After solution treating at 410 °C the specimen was quenched to a temperature of 220 °C for 20 min followed by 90 s at 277 °C and finally water quenched. It is apparent that the mean interlamellar spacing is higher at higher ageing temperatures. As with eutectic solidification this is because less free energy is available for the formation of α/β interfaces when the total driving force for transformation is reduced.

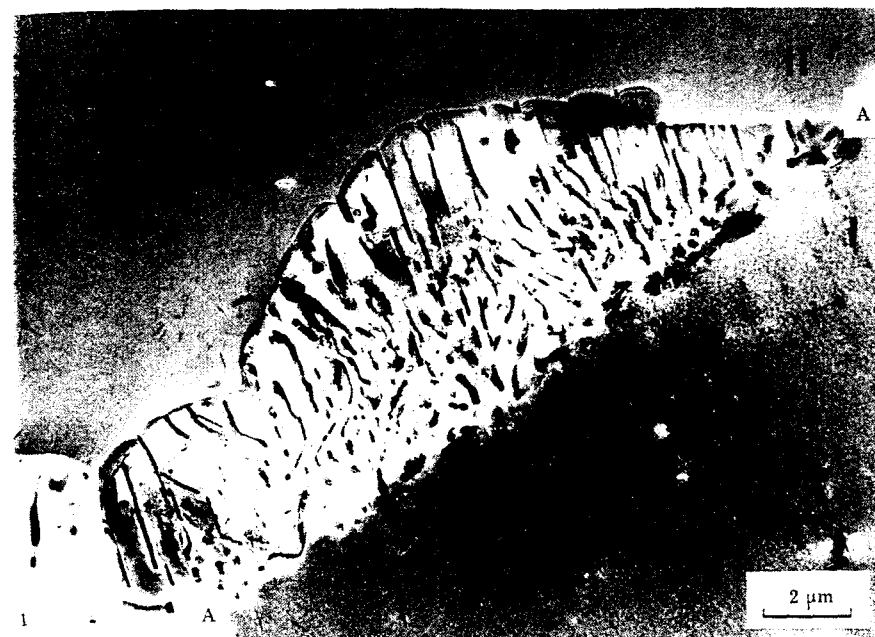


Fig. 5.51 Cellular precipitation of $\text{Mg}_{17}\text{Al}_{12}$ in an Mg–9 at% Al alloy solution treated and aged 1 h at 220 °C followed by 2 min at 310 °C. Some general $\text{Mg}_{17}\text{Al}_{12}$ precipitation has also occurred on dislocations within the grains.

The growth of cellular precipitates requires the partitioning of solute to the tips of the precipitates in contact with the advancing grain boundary. This can occur in one of two ways: either by diffusion through the lattice ahead of the advancing cell front, or by diffusion in the moving boundary. Partitioning by lattice diffusion would require solute concentration gradients ahead of the cell front while, if the grain boundary is the most effective diffusion route, the matrix composition should remain unchanged right up to the cell front. In the case of the Mg–Al alloy it has been possible to do microanalysis with sufficiently high spatial resolution to resolve these possibilities directly. (The technique used was electron energy loss spectroscopy using plasmon losses¹³.) The results of such measurements, Fig. 5.54a, clearly indicate that the matrix composition remains unchanged to within 10 nm of the advancing cell front so that partitioning must be taking place within the boundary itself. This is to be expected since precipitation is occurring at relatively low temperatures where solute transport tends to become more effective via grain boundaries than through the lattice.

Figure 5.54b shows the aluminium concentration in the α matrix along a line between the β ($\text{Mg}_{17}\text{Al}_{12}$) lamellae. This is essentially a replica of a similar concentration profile that must exist within the advancing grain boundary. Therefore apart from the matrix in contact with the β precipitate, the cell matrix is still supersaturated with respect to equilibrium.

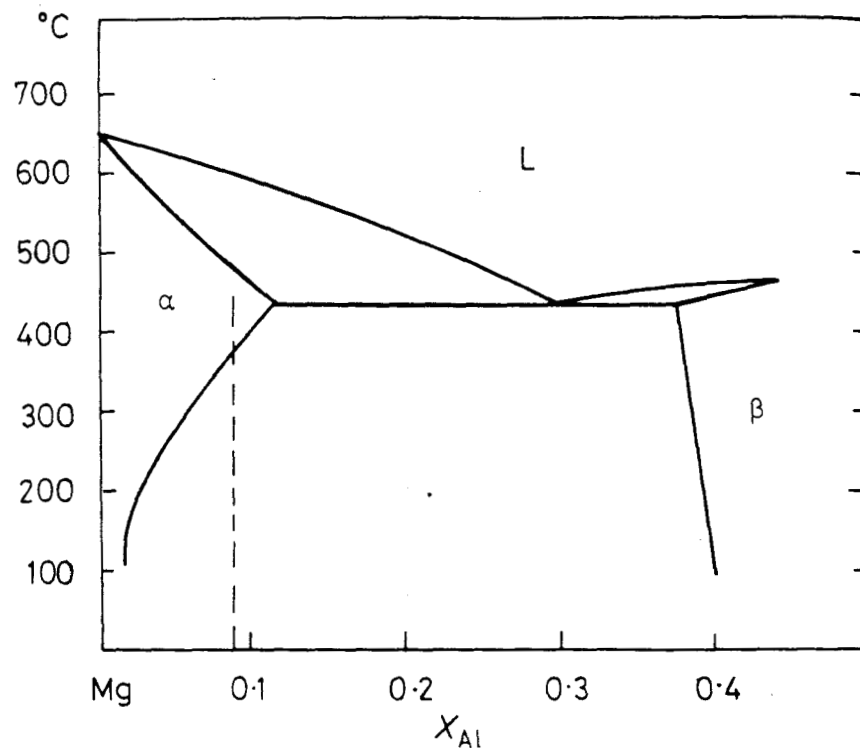


Fig. 5.52 The relevant part of the Mg-Al phase diagram.

Cellular precipitation is also known as *discontinuous precipitation* because the composition of the matrix changes discontinuously as the cell front passes. Precipitation that is not cellular is referred to as general or continuous because it occurs generally throughout the matrix on dislocations or grain boundaries, etc. and the matrix composition at a given point decreases continuously with time. Often general precipitation leads to a finely distributed intermediate precipitate that is associated with good mechanical properties. The cellular reaction is then unwanted because the intermediate precipitates will dissolve as they are overgrown and replaced by the coarse equilibrium precipitates within the cells.

5.8 Eutectoid Transformations

5.8.1 The Pearlite Reaction in Fe-C Alloys

When austenite containing about 0.8wt% C is cooled below the A_1 temperature it becomes simultaneously supersaturated with respect to ferrite and cementite and a eutectoid transformation results, i.e.

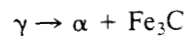


Fig. 5.53 A cell formed during ageing at two temperatures: 30 min at 220 °C followed by 30 min at 277 °C and water quenched. Note the change in interlamellar spacing caused by the change in undercooling.

The manner in which this reaction occurs is very similar to a eutectic transformation where the original phase is a liquid instead of a solid. In the case of Fe-C alloys the resultant microstructure comprises lamellae, or sheets, of cementite embedded in ferrite as shown in Fig. 5.55. This is known as pearlite. Both cementite and ferrite form directly in contact with the austenite as shown.

Pearlite nodules nucleate on grain boundaries and grow with a roughly constant radial velocity into the surrounding austenite grains. At small undercoolings below A_1 the number of pearlite nodules that nucleate is relatively small, and the nodules can grow as hemispheres or spheres without interfering with each other. At larger undercoolings the nucleation rate is much higher and *site saturation* occurs, that is all boundaries become quickly covered with nodules which grow together forming layers of pearlite outlining the prior austenite grain boundaries, Fig. 5.56.

Nucleation of Pearlite

The first stage in the formation of pearlite is the nucleation of either cementite or ferrite on an austenite grain boundary. Which phase nucleates first will depend on the grain-boundary structure and composition. Suppose that it is cementite. The cementite will try to minimize the activation energy barrier to

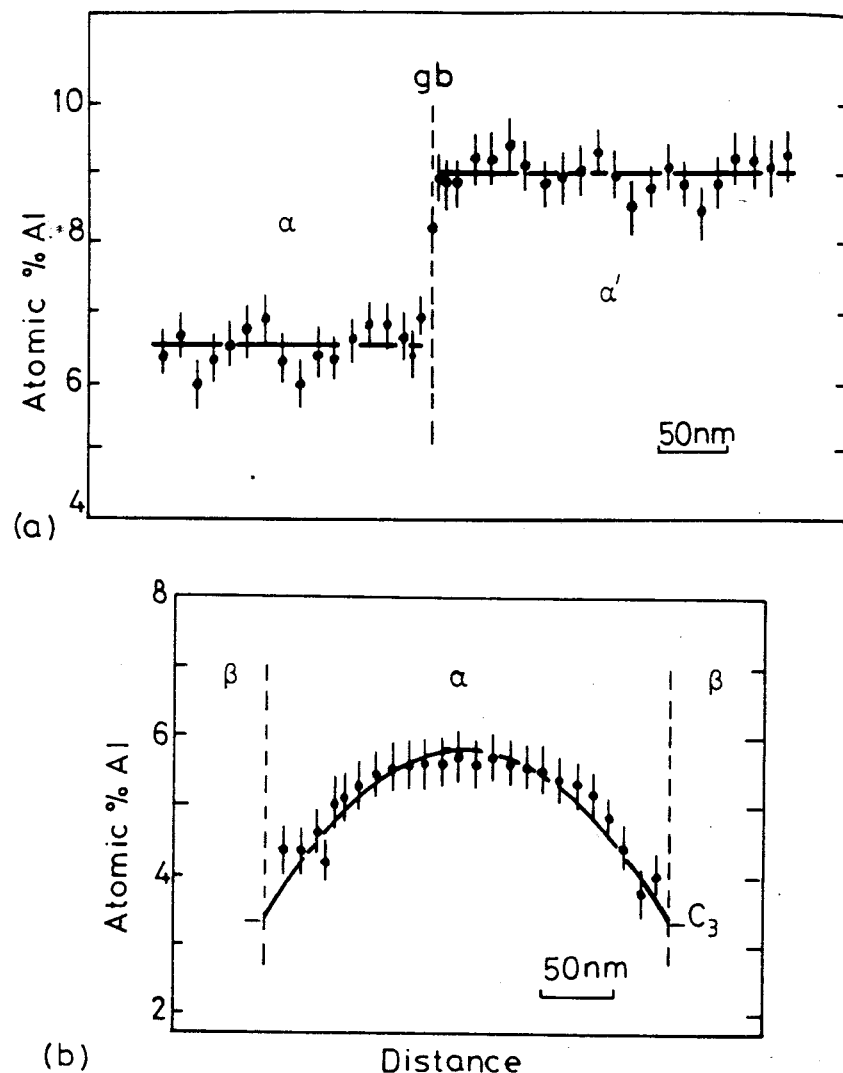


Fig. 5.54 (a) The variation of aluminium concentration across an advancing grain boundary midway between two precipitate lamellae. (b) A similar profile along a line such as S in Fig. 5.53.

nucleation by forming with an orientation relationship to one of the austenite grains, γ_1 in Fig. 5.57a. (The crystal structure of cementite is orthorhombic and the orientation relationship is close to $(100)_c // (111)_\gamma$, $(010)_c // (110)_\gamma$, $(001)_c // (\bar{1}12)_\gamma$.) Therefore the nucleus will have a semicoherent, low-mobility interface with γ_1 and an incoherent mobile interface with γ_2 . The austenite surrounding this nucleus will become depleted of carbon which will increase the driving force for the precipitation of ferrite, and a ferrite nucleus forms adjacent to the cementite nucleus also with an orientation relationship to γ_1 (the Kurdjumov-Sachs relationship). This process can be repeated



Fig. 5.55 A pearlite colony advancing into an austenite grain. (After L.S. Darken and R.M. Fisher in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), by permission of The Metallurgical Society of AIME.

cleation of both phases the colony can grow edgewise by the movement of the *incoherent* interfaces, that is pearlite grows into the austenite grain with which it does not have an orientation relationship. The carbon rejected from the growing ferrite diffuses through the austenite to in front of the cementite, as with eutectic solidification.

If the alloy composition does not perfectly correspond to the eutectoid composition the grain boundaries may already be covered with a proeutectoid ferrite or cementite phase. If, for example, the grain boundary already contains a layer of cementite, the first ferrite nucleus will form with an orientation relationship to this cementite on the mobile incoherent side of the allotriomorphs as shown in Fig. 5.57b. Again due to the higher mobility of the incoherent interfaces the pearlite will grow into the austenite with which there is no orientation relationship.

Whatever the pearlite nucleation mechanism, new cementite lamellae are able to form by the branching of a single lamella into two new lamellae as shown in Fig. 5.57a(iv) or c. The resultant pearlite colony is effectively two interpenetrating single crystals.

It can be seen that the nucleation of pearlite requires the establishment of cooperative growth of the two phases. It takes time for this cooperation to be established and the rate of colony nucleation therefore increases with time. In



Fig. 5.56 A partially transformed eutectoid steel. Pearlite has nucleated on grain boundaries and inclusions ($\times 100$). (After J.W. Cahn and W.C. Hagel in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), 1962, by permission of The Metallurgical Society of AIME.)

some cases cooperation is not established and the ferrite and cementite grow in a non-lamellar manner producing so-called degenerate pearlite¹⁴.

Pearlite Growth

The growth of pearlite in binary Fe–C alloys is analogous to the growth of a lamellar eutectic with austenite replacing the liquid. Carbon can diffuse interstitially through the austenite to the tips of the advancing cementite lamellae so that the equations developed in Section 4.3.2 should apply equally well to pearlite. Consequently the minimum possible interlamellar spacing (S^*) should vary inversely with undercooling below the eutectoid temperature (A_1), and assuming the observed spacing (S_0) is proportional to S^* gives

$$S_0 \propto S^* \propto (\Delta T)^{-1} \quad (5.53)$$

Similarly the growth rate of pearlite colonies should be constant and given by a relationship of the type

$$v = kD_c^{\frac{1}{2}}(\Delta T)^2 \quad (5.54)$$

where k is a thermodynamic term which is roughly constant.

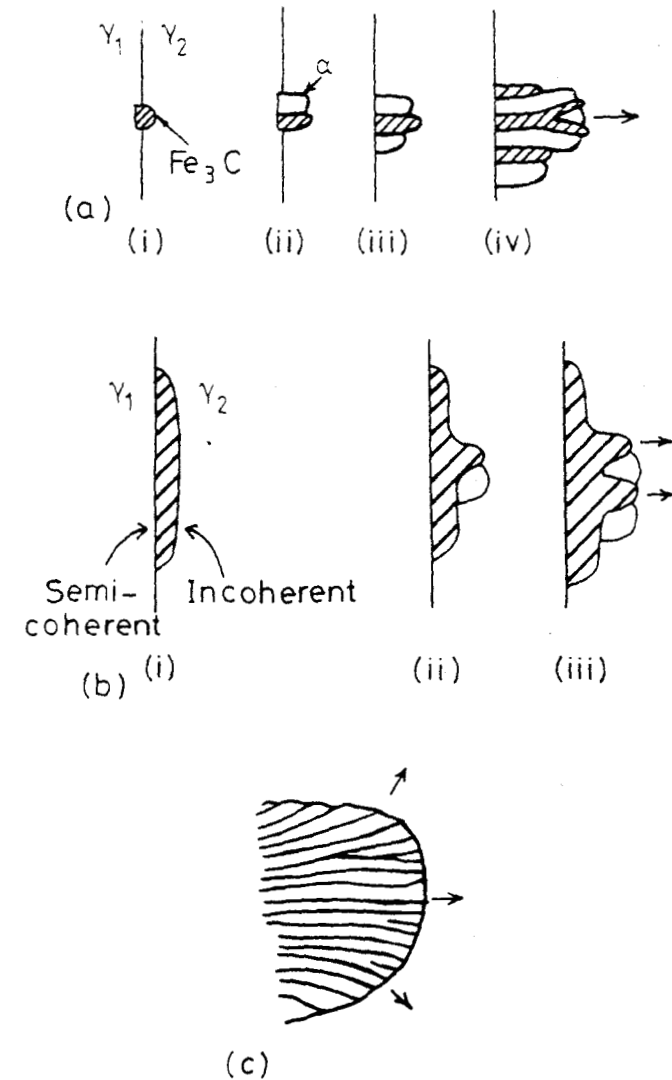


Fig. 5.57 Nucleation and growth of pearlite. (a) On a 'clean' grain boundary. (i) Cementite nucleates on grain boundary with coherent interface and orientation relationship with γ_1 and incoherent interface with γ_2 . (ii) α nucleates adjacent to cementite also with a coherent interface and orientation relationship with γ_1 . (This also produces an orientation relationship between the cementite and ferrite.) (iii) The nucleation process repeats sideways, while *incoherent* interfaces grow into γ_2 . (iv) New plates can also form by a branching mechanism. (b) When a proeutectoid phase (cementite or ferrite) already exists on that boundary, pearlite will nucleate and grow on the incoherent side. A different orientation relationship between the cementite and ferrite results in this case. (c) A pearlite colony at a later stage of growth.

Observed spacings are found to obey Equation 5.53, varying from $\sim 1 \mu\text{m}$ at high temperatures to $\sim 0.1 \mu\text{m}$ at the lowest temperatures of growth¹⁵. However, it is found that S_0 is usually greater than $2S^*$, i.e. the observed spacing is not determined by the maximum growth rate criterion. Instead it may be determined by the need to create new cementite lamellae as the perimeter of the pearlite nodules increases. This can occur either by the nucleation of new cementite lamellae, or by the branching of existing lamellae, Fig. 5.57c.

In the case of binary Fe-C alloys, observed growth rates are found to agree rather well with the assumption that the growth velocity is controlled by the diffusion of carbon in the austenite. Figure 5.58 shows measured and calculated growth rates as a function of temperature. The calculated line is based on an equation similar to Equation 5.54 and shows that the measured growth rates are reasonably consistent with volume-diffusion control. However, it is also possible that some carbon diffusion takes place through the γ/α and $\gamma/\text{cementite}$ interfaces, which could account for the fact that the predicted growth rates shown in Fig. 5.58 are consistently too low.

A schematic TTT diagram for the pearlite reaction in eutectoid Fe-C alloys is shown in Fig. 5.59. Note the 'C' shape typical of diffusional transformations that occur on cooling. The maximum rate of transformation occurs at about 550°C . At lower temperatures another type of transformation product, namely *Bainite*, can grow faster than pearlite. This transformation is dealt with in the next section.

Eutectoid transformations are found in many alloys besides Fe-C. In alloys where all elements are in substitutional solid solution, lattice diffusion is found to be too slow to account for observed growth rates. In these cases diffusion

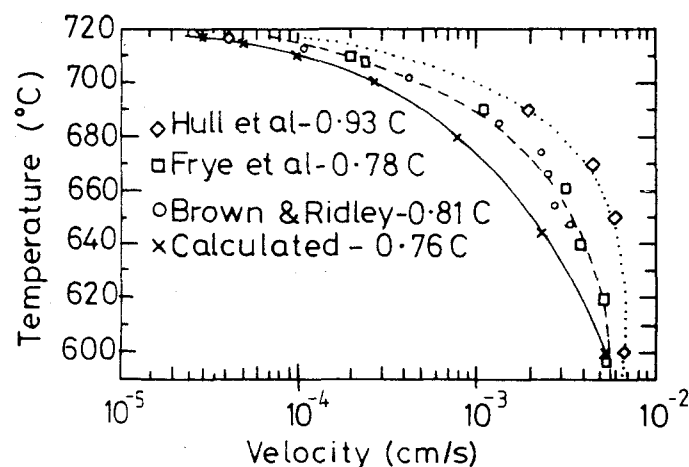


Fig. 5.58 Pearlite growth rate v , temperature for plain carbon steels. (After M.P. Puls and J.S. Kirkaldy, *Metallurgical Transactions* 3 (1972) 2777. © American Society for Metals and the Metallurgical Society of AIME, 1972.)

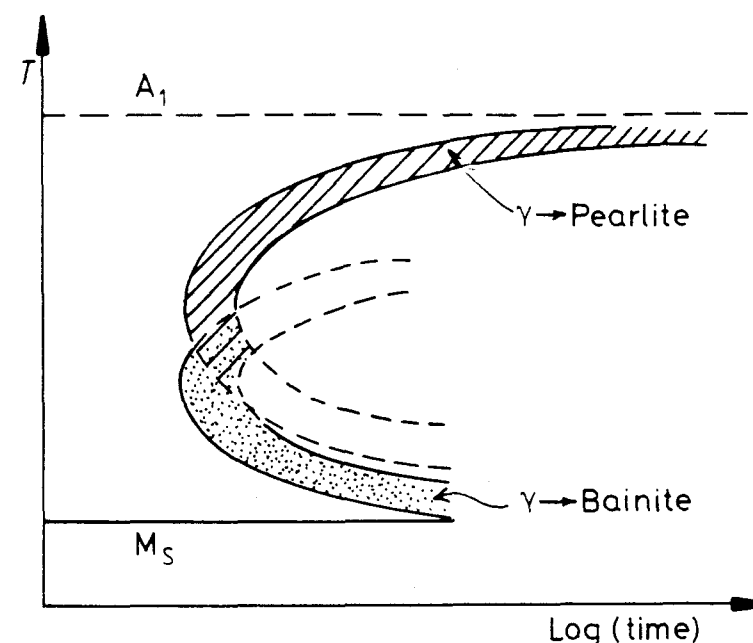


Fig. 5.59 Schematic diagram showing relative positions of the transformation curves for pearlite and bainite in plain carbon eutectoid steels.

occurs instead through the colony/matrix interface. Consideration of the diffusion problem in this case leads to a relationship of the type

$$v = kD_B(\Delta T)^3 \quad (5.55)$$

where k is a thermodynamic constant and D_B is the boundary diffusion coefficient.

Pearlite in Off-Eutectoid Fe-C Alloys

When austenite containing more or less carbon than the eutectoid composition is isothermally transformed below the A_1 temperature the formation of pearlite is usually preceded by the precipitation of proeutectoid ferrite or cementite. However, if the undercooling is large enough and the departure from the eutectoid composition is not too great it is possible for austenite of non-eutectoid composition to transform directly to pearlite. The region in which this is possible corresponds approximately to the condition that the austenite is simultaneously saturated with respect to both cementite and ferrite, i.e. the hatched region in Fig. 5.60. (See also Fig. 5.48). Thus a 0.6% C alloy, for example can be transformed to $\sim 100\%$ pearlite provided the temperature is low enough to bring the austenite into the hatched region of Fig. 5.60 (but not so low that bainite forms). At intermediate undercoolings some proeutectoid ferrite will form but less than predicted by the equilibrium phase diagram.

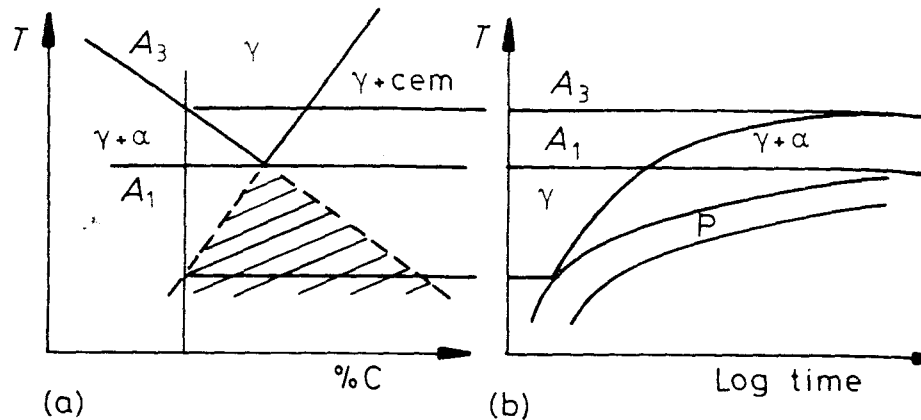


Fig. 5.60 Effect of transformation temperature on the volume fraction of proeutectoid ferrite.

Similar considerations apply to transformations during continuous cooling—larger grain sizes and faster cooling rates favour low volume fractions of ferrite. Compare Fig. 5.49c and d.

5.8.2 The Bainite Transformation

When austenite is cooled to large supersaturations below the nose of the pearlite transformation curve a new eutectoid product called *bainite* is produced. Like pearlite, bainite is a mixture of ferrite and carbide, but it is microstructurally quite distinct from pearlite and can be characterized by its own C curve on a TTT diagram. In plain carbon steels this curve overlaps with the pearlite curve (Fig. 5.59) so that at temperatures around 500 °C both pearlite and bainite form competitively. In some alloy steels, however, the two curves are separated as shown in Fig. 5.65.

The microstructure of bainite depends mainly on the temperature at which it forms¹⁶.

Upper Bainite

At high temperatures (350 °C–550 °C) bainite consists of needles or laths of ferrite with cementite precipitates between the laths as shown in Fig. 5.61. This is known as *upper bainite*. Figure 5.61a shows the ferrite laths growing into partially transformed austenite. The light contrast is due to the cementite. Figure 5.61b illustrates schematically how this microstructure is thought to develop. The ferrite laths grow into the austenite in a similar way to Widmanstätten side-plates. The ferrite nucleates on a grain boundary with a Kurdjumov–Sachs orientation relationship with one of the austenite grains, γ_2 , say. Since the undercooling is very large the nucleus grows most rapidly into the γ_2 grain forming ferrite laths with low energy semicoherent inter-

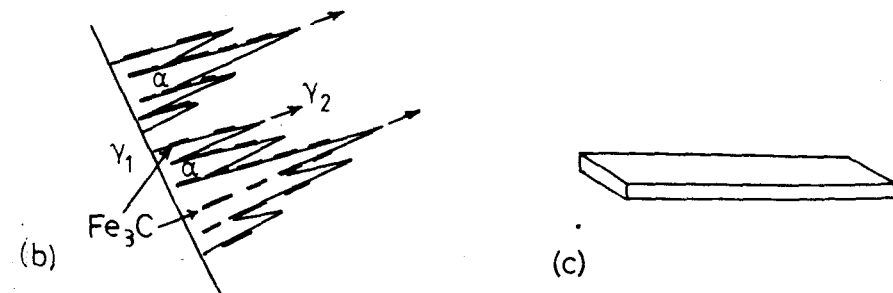
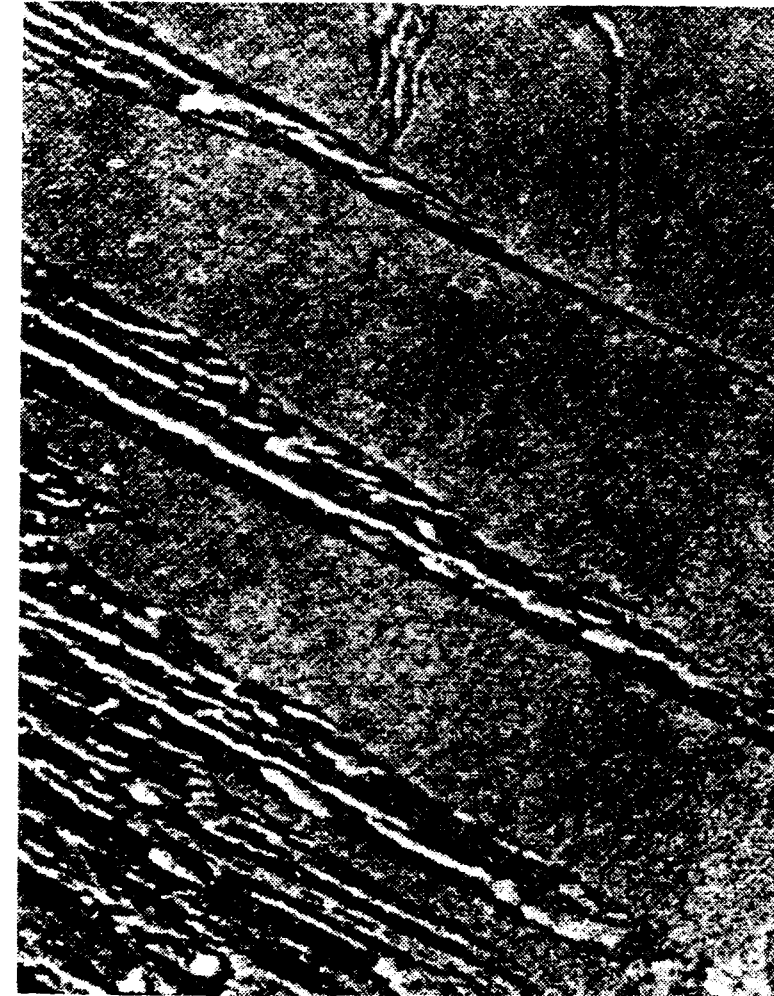


Fig. 5.61 (a) Upper bainite in medium-carbon steel (replica $\times 13\,000$) (by permission of the Metals Society). (b) Schematic of growth mechanism. Widmanstätten ferrite laths growth into γ_2 . (α and γ_2 have Kurdjumov–Sachs orientation relationship.) Cementite plates nucleate in carbon-enriched austenite. (c) Illustrating the shape of a 'lath'.

faces. This takes place at several sites along the boundary so that a group of finely spaced laths develops. As the laths thicken the carbon content of the austenite increases and finally reaches such a level that cementite nucleates and grows.

At the higher temperatures of formation upper bainite closely resembles finely spaced Widmanstätten side-plates, Fig. 5.46d. As the temperature decreases the bainitic laths become narrower so that individual laths may only be resolved by electron microscopy.

At the highest temperatures where pearlite and bainite grow competitively in the same specimen it can be difficult to distinguish the pearlite colonies from the upper bainite. Both appear as alternate layers of cementite in ferrite. The discontinuous nature of the bainitic carbides does not reveal the difference since pearlitic cementite can also appear as broken lamellae. However, the two microstructures have formed in quite different ways. The greatest difference between the two constituents lies in their crystallography. In the case of pearlite the cementite and ferrite have no specific orientation relationship to the austenite grain in which they are growing, whereas the cementite and ferrite in bainite do have an orientation relationship with the grain in which they are growing. This point is illustrated in Fig. 5.62. The micrograph is from a hypoeutectoid steel (0.6% C) which has been partially transformed at 710 °C and then quenched to room temperature, whereupon the untransformed austenite was converted into martensite. The quench, however, was not fast enough to prevent further transformation at the γ/α interface. The dark constituent is very fine pearlite which was nucleated on the *incoherent* α/γ interface, across which there is no orientation relationship. The ferrite and lower austenite grain, however, have an orientation relationship which has led to bainite formation.

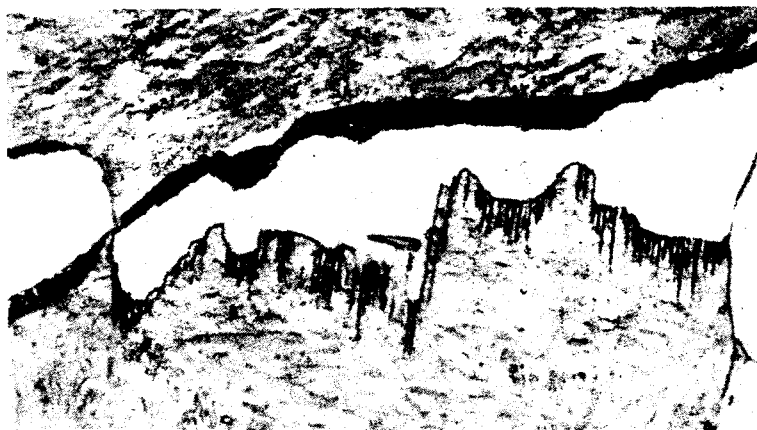


Fig. 5.62 Hypoeutectoid steel (0.6% C) partially transformed for 30 min at 710 °C, inefficiently quenched. Bainitic growth into lower grain of austenite and pearlitic growth into upper grain during quench ($\times 1800$). (After M. Hillert in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), 1962, by permission of the Metallurgical Society of AIME.)

Lower Bainite

At sufficiently low temperatures the microstructure of bainite changes from laths into plates and the carbide dispersion becomes much finer, rather like in tempered martensite. The temperature at which the transition to lower bainite occurs depends on the carbon content in a complex manner. For carbon levels below about 0.5wt% the transition temperature increases with increasing carbon, from 0.5–0.7wt% C it decreases and above approximately 0.7wt% C it is constant at about 350 °C. At the temperatures where lower bainite forms the diffusion of carbon is slow, especially in the austenite and carbides precipitate in the ferrite with an orientation relationship. The carbides are either cementite or metastable transition carbides such as ϵ -carbide and they are aligned at approximately the same angle to the plane of the ferrite plate (Fig. 5.63). The habit plane of the ferrite plates in lower bainite is the same as that of the martensite that forms at lower temperatures in the same alloy. As with upper bainite, some carbides can also be found between the ferrite plates.

The different modes of formation of upper and lower bainite result in different transformation kinetics and separate C curves on the TTT diagram. An example, the case of a low-alloy steel, is shown in Fig. 5.68.

Transformation Shears

If a polished specimen of austenite is transformed to bainite (upper or lower) it is found that the growth of bainite laths or plates produces a surface relief effect like that of martensite plates. For example Fig. 5.64 shows the surface tilts that result from the growth of lower bainite plates. This has been interpreted as suggesting that the bainite plates form by a shear mechanism in the same way as the growth of martensite plates (see Chapter 6). In other words it is supposed that the iron atoms are transferred across the ferrite/austenite interface in an ordered military manner. However, the growth rate of the bainite plates is controlled by the rate at which carbon can diffuse away from the interface, or by the rate at which carbides can precipitate behind the interface, whereas martensite plates are able to advance without any carbon diffusion, and the plates can grow as fast as the glissile interfaces can advance.

There is, however, much uncertainty regarding the mechanism by which bainitic ferrite grows, and the nature of the austenite–ferrite interface in martensite and bainite. In fact the formation of Widmanstätten side-plates also leads to surface tilts of the type produced by a shear transformation. Also the phenomenological theory of martensite is able to account for the observed orientation relationships and habit planes found in Widmanstätten plates as well as bainite and martensite. It can be seen, therefore, that some phase transformations are not exclusively military or civilian, but show characteristics common to both types of transformation. For a detailed review of the bainite transformation the reader should consult the article by Bhadeshia and Christian, given in the Further Reading section at the end of this chapter.

5.8.3 The Effect of Alloying Elements on Hardenability

The primary aim of adding alloying elements to steels is to increase the hardenability, that is, to delay the time required for the decomposition into

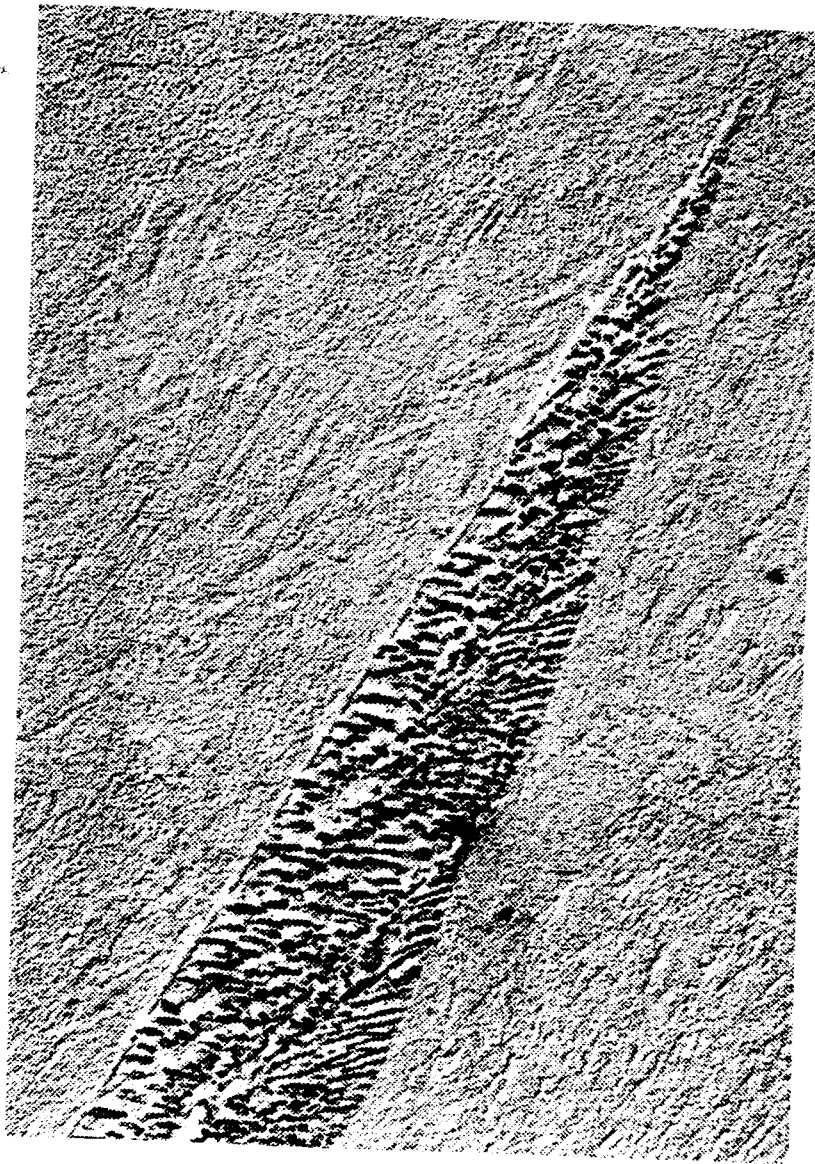
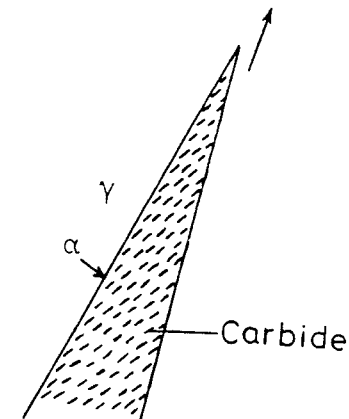


Fig. 5.63 (a) Lower bainite in 0.69wt% C low-alloy steel (replica $\times 1100$). (After R.F. Heheman in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 196.) (b) A possible growth mechanism. α/γ interface advances as fast as carbides precipitate at interface thereby removing the excess carbon in front of the α .



(b)

Fig. 5.63 (continued)

ferrite and pearlite. This allows slower cooling rates to produce fully martensitic structures. Figure 5.65 shows some examples of TTT diagrams for various low-alloy steels containing Mn, Cr, Mo and Ni in various combinations and concentrations. Note the appearance of two separate C curves for pearlite and bainite, and the increasing time for transformation as the alloy content increases.

Basically there are two ways in which alloying elements can reduce the rate of austenite decomposition. They can reduce either the growth rate or the nucleation rate of ferrite, pearlite, or bainite.

The main factor limiting hardenability is the rate of formation of pearlite at the nose of the C curve in the TTT diagram. To discuss the effects of alloy elements on pearlite growth it is necessary to distinguish between austenite stabilizers (e.g. Mn, Ni, Cu) and ferrite stabilizers (e.g. Cr, Mo, Si). Austenite stabilizers depress the A_1 temperature, while ferrite stabilizers have the opposite effect. All of these elements are substitutionally dissolved in the austenite and ferrite.

At equilibrium an alloy element X will have different concentrations in cementite and ferrite, i.e. it will partition between the two phases. Carbide-forming elements such as Cr, Mo, Mn will concentrate in the carbide while elements like Si will concentrate in the ferrite. When pearlite forms close to the A_1 temperature the driving force for growth will only be positive if the equilibrium partitioning occurs. Since X will be homogeneously distributed within the austenite, the pearlite will only be able to grow as fast as substitutional diffusion of X allows partitioning to occur. The most likely diffusion route for substitutional elements is through the γ/α and $\gamma/\text{cementite}$ interfaces. However, it will be much slower than the interstitial diffusion of carbon and will therefore reduce the pearlite growth rate.

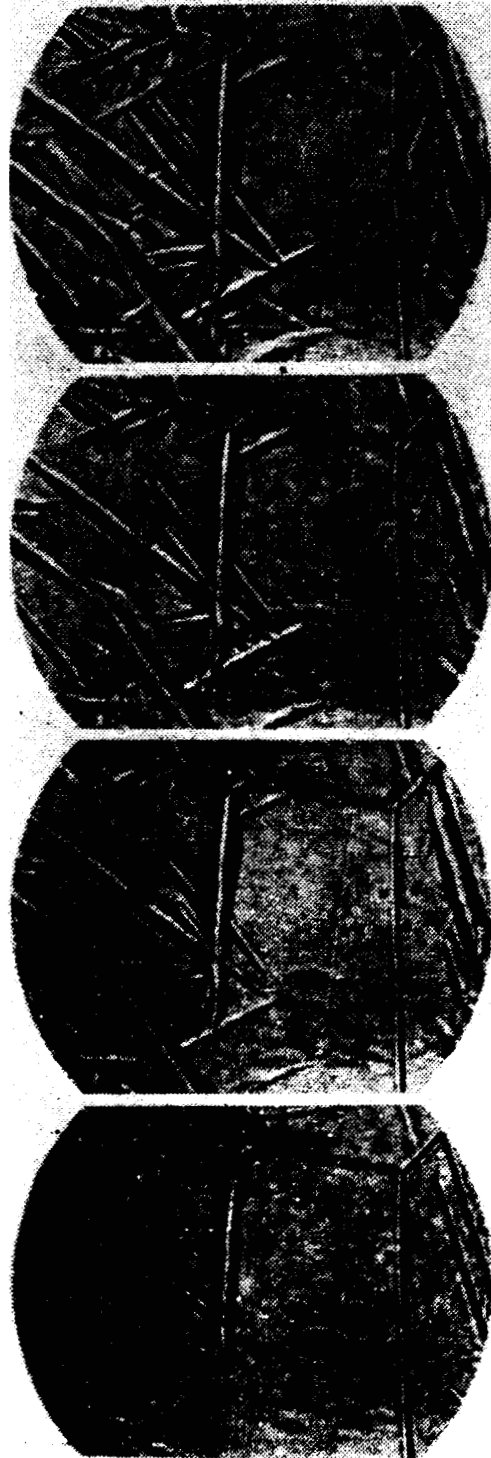


Fig. 5.64 Photos taken with a hot-stage microscope showing the nucleation and growth of bainite plates at 350 °C. The contrast is due to surface relief. (a) 14.75 min; (b) 16.2 min; (c) 17.25 min; (d) 19.2 min. (After G.R. Speich, in *Decomposition of Austenite by Diffusional Processes*, V.F. Zackay and H.I. Aaronson (Eds.), 1962, by permission of The Metallurgical Society of AIME.)

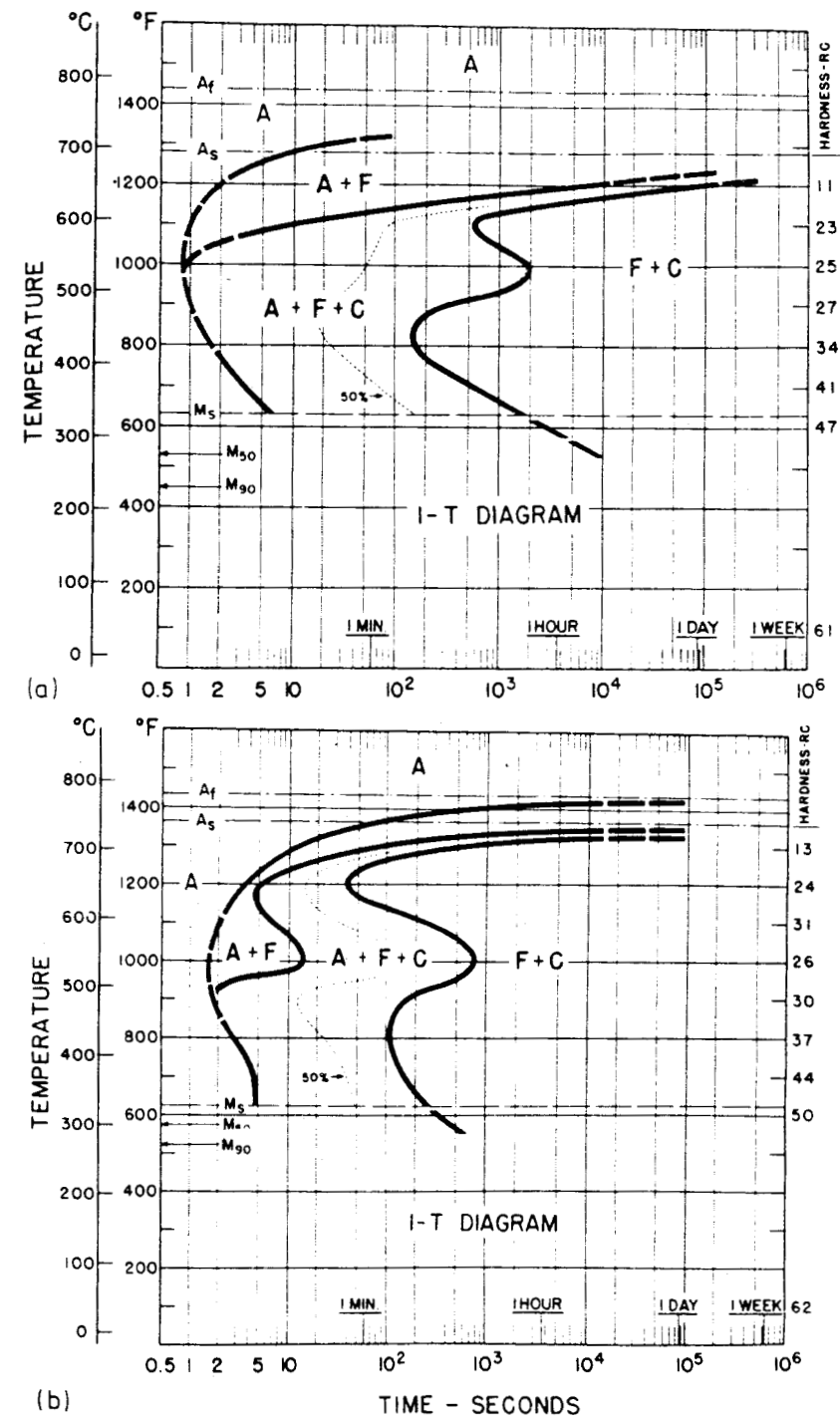
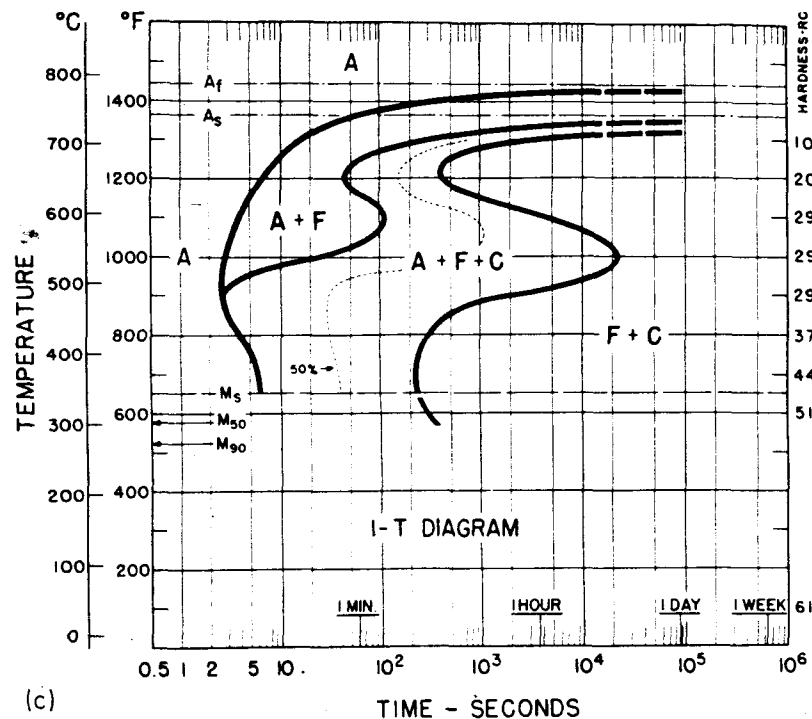
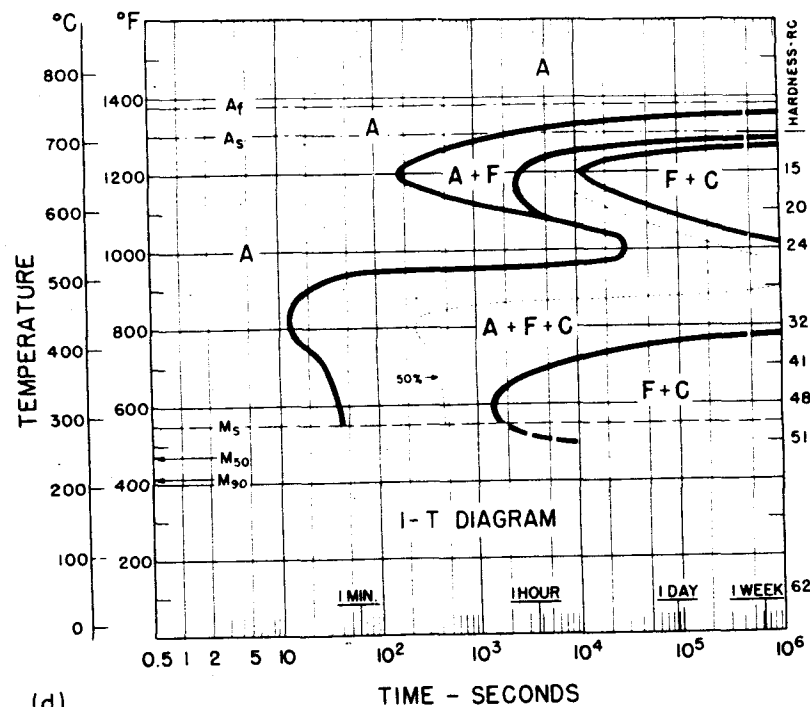


Fig. 5.65 TTT diagrams for four commercial low-alloy steels all of which contain roughly 0.4% C and 1% Mn. In addition (b) contains 0.9% Cr. (c) contains 1.0% Cr and 0.2% Mo, and (d) contains 0.8% Cr, 0.3% Mo, and 1.8% Ni. Note the tendency to form two distinct knees, one for pearlite formation and one for bainite formation. (From *Atlas of Isothermal Transformation and Cooling Transformation Diagrams*, American Society for Metals, 1977.)



(c)



(d)

When X is a ferrite stabilizer there are thermodynamic considerations that suggest that X will partition even at large undercoolings close to the nose of the C curve. Thus Si, for example, will increase the hardenability by diffusing along the austenite/pearlite interface into the ferrite. The partitioning of alloying elements in an Fe-0.6 wt% C-0.85% Cr-0.66% Mn-0.26% Si steel transformed at 597 °C for 2 min is shown in Fig. 5.66.

When X is an austenite stabilizer such as Ni, it is possible, at sufficiently high undercoolings, for pearlite to grow without partitioning. The ferrite and cementite simply inherit the Ni content of the austenite and there is no need for substitutional diffusion. Pearlite can then grow as fast as diffusion of carbon allows. However, the growth rate will still be lower than in binary Fe-C alloys since the non-equilibrium concentration of X in the ferrite and cementite will raise their free energies, thereby lowering the eutectoid temperature, Fig. 5.67, and reducing the total driving force. For the same reasons zero-partitioning is only possible at temperatures below the metastable eutectoid as shown in Fig. 5.67.

When X is a strong carbide-forming element such as Mo or Cr it has been suggested¹⁷ that it can reduce the rate of growth of pearlite, as well as proeutectoid ferrite, by a solute-drag effect on the moving γ/α interface. These elements also partition to cementite as shown in Fig. 5.66.

Hardenability is not solely due to growth-rate effects. It is also possible that the alloying elements affect the rate of nucleation of cementite or ferrite. For

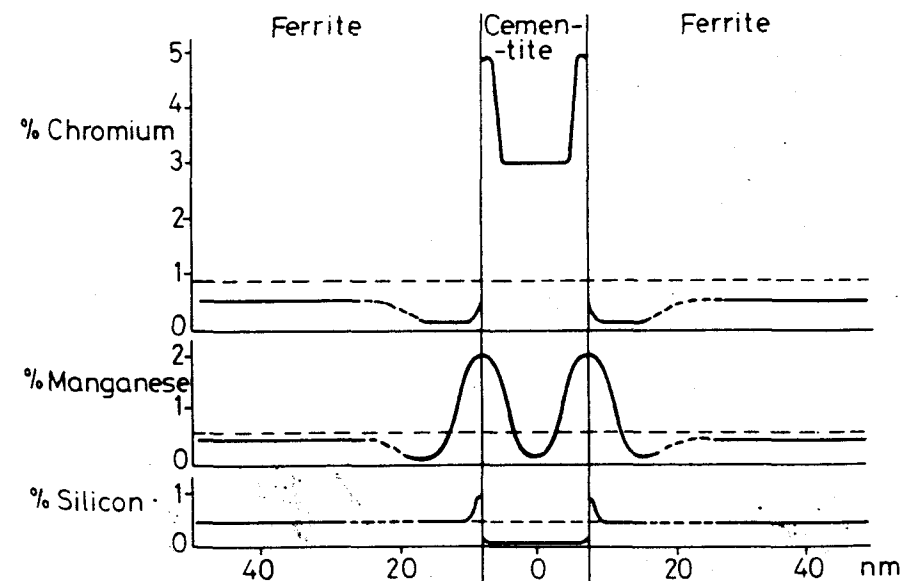


Fig. 5.66 Schematic diagram showing the measured variations of alloying elements in pearlite. These measurements were made using a time-of-flight atom probe. (P.R. Williams, M.K. Miller, P.A. Beavan and G.D.W. Smith, *Phase Transformations*, Vol. 2, Institute of Metallurgists, 1979, p. 98.)

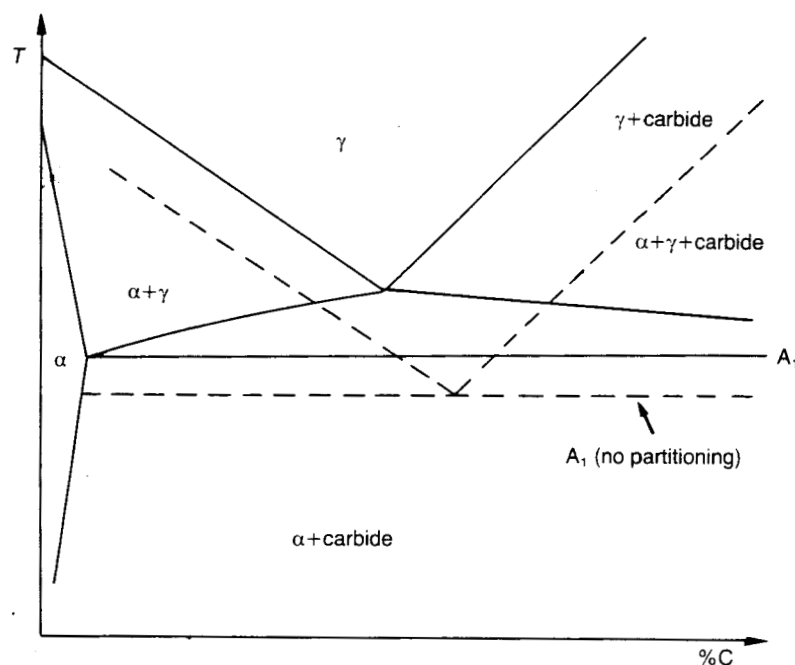


Fig. 5.67 Schematic phase diagram for Fe-C-X alloy where X is a substitutional element. Between the solid and dashed lines, precipitation can occur in the austenite only if X is partitioned between the phases.

example, it has been suggested¹⁸ that the 'bay' at $\sim 500^\circ\text{C}$ in the TTT diagrams of steels containing Cr, Mo and B (see Fig. 5.65) may be due to the poisoning of ferrite nucleation sites by the precipitation of X-carbide clusters in grain boundaries.

The diagrams shown in Fig. 5.65 are not entirely accurate especially with regard to the bainite transformation at temperatures in the vicinity of the M_s temperatures. It has been found that below the M_s temperature the bainite transformation rate is greatly increased by the martensite-transformation strains. The TTT diagram for the bainite transformation in Fig. 5.65d has recently been redetermined using a new experimental technique based on magnetic permeability measurements¹⁹ and the results are shown in Fig. 5.68. The acceleration of the transformation close to M_s and the existence of separate C curves for upper and lower bainite are apparent.

5.8.4 Continuous Cooling Diagrams

Isothermal transformation (TTT) diagrams are obtained by rapidly quenching to a given temperature and then measuring the volume fraction of the various

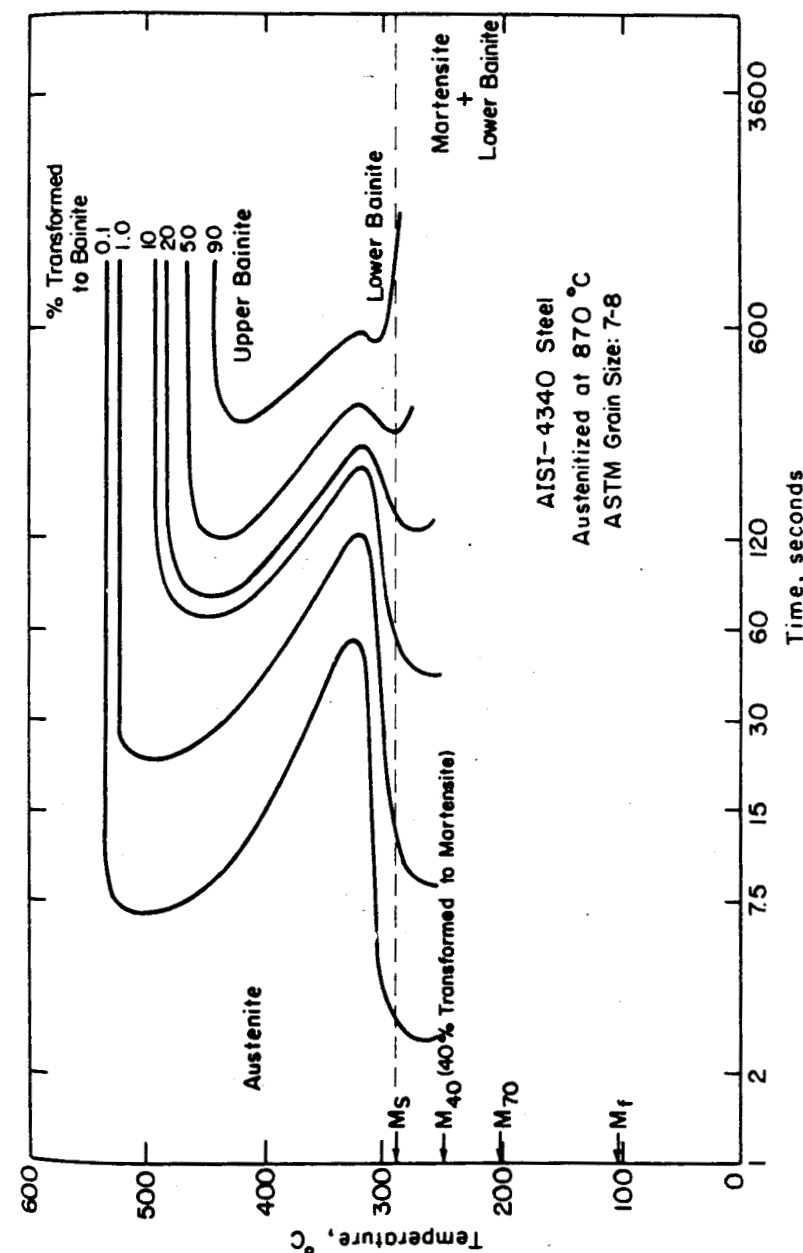


Fig. 5.68 TTT diagrams for steel in Fig. 5.65(d) determined by a magnetic permeability technique. (After V.F. Zackay and E.R. Parker, in *Alloy and Microstructural Design*, J.K. Tien and G.S. Ansell (Eds.), Academic Press, New York, 1976.)

constituents that form as a function of time at that temperature. Practical heat treatments, however, are usually concerned with transformations that occur during continuous cooling, and under these conditions TTT diagrams cannot be used to give the times and temperatures of the various transformations. A continuous cooling transformation (CCT) diagram must be used instead.

To a first approximation the CCT diagram is the TTT diagram shifted to lower temperatures and longer times. This can be understood as follows. In a specimen held at a constant temperature the transformation starts when the product (Nt) reaches a certain value, α , say. In a continuously cooled sample, time near the start of cooling is not very effective since N is low at low supercoolings. Therefore when the cooling curve reaches the TTT start curve the total value of Nt will be less than α and further time (and therefore cooling) will be required before the start of the CCT diagram is reached. Similarly the end of the reaction will be displaced to lower temperatures and longer times. The relationship between a CCT and an TTT diagram for a eutectoid steel is shown in Fig. 5.69. Note that whereas the TTT diagram is interpreted by reading from left to right at a constant temperature the CCT diagram is read along the cooling curves from the top left to bottom right. The cooling curves in Fig. 5.69 refer to various distances from the quenched end of a Jominy end-quench specimen. Transformation occurs along the hatched parts of the lines. Figure 5.69 is in fact simplified and cooling along B would lead to the production of some bainite. But otherwise it can be seen that point B will transform partly to fine pearlite at high temperatures around 500–450 °C. Between 450 and 200 °C the remaining austenite will be unable to transform and below 200 °C transformation to martensite occurs.

The above relationship between TTT and CCT diagrams is only approximate. There are several features of CCT diagrams that have no counterpart in TTT diagrams especially in alloy steels. These include the following: (i) a depression of the M_s temperature at slow cooling rates, (ii) the tempering of martensite that takes place on cooling from M_s to about 200 °C, (iii) a greater variety of microstructures.

Figure 5.70 shows more complete CCT diagrams for a medium-carbon steel with different Mn contents. These diagrams were obtained with a high-speed dilatometer using programmed linear cooling rates for all except the highest quench rates. For each cooling curve the cooling rate and volume fractions of ferrite and pearlite are indicated. Note how the volume fraction of pearlite increases as the cooling rate is increased from 2.5 to 2300 °F/min in the low-Mn steel. In practical heat treatments the cooling curves will not be linear but will depend on the transfer of heat from the specimen to the quenching medium and the rate of release of latent heat during transformation. In general, the evolution of latent heat reduces the rate of cooling during the transformation range and can even lead to a rise in temperature, i.e. recalescence. Recalescence is often associated with the pearlite transformation when the growth rate is very high, e.g. in unalloyed steels, but the effect can also be seen quite clearly for a cooling rate of 4100 °F/min in Fig. 5.70a.

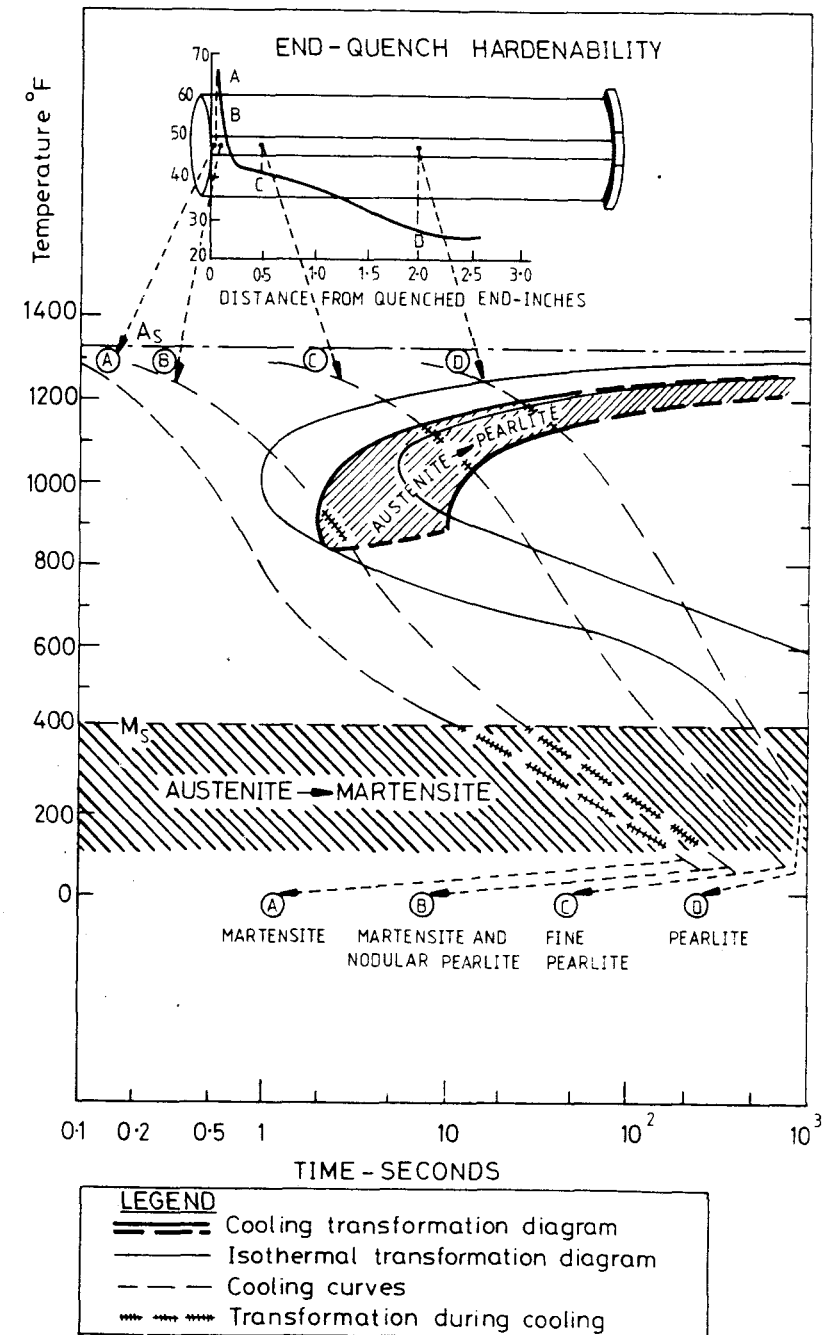


Fig. 5.69 Correlation of continuous cooling and isothermal transformations with end-quench hardenability test data for eutectoid carbon steel. (*Atlas of Isothermal Transformation and Cooling Transformation Diagrams*, American Society for Metals, 1977, p. 376.)

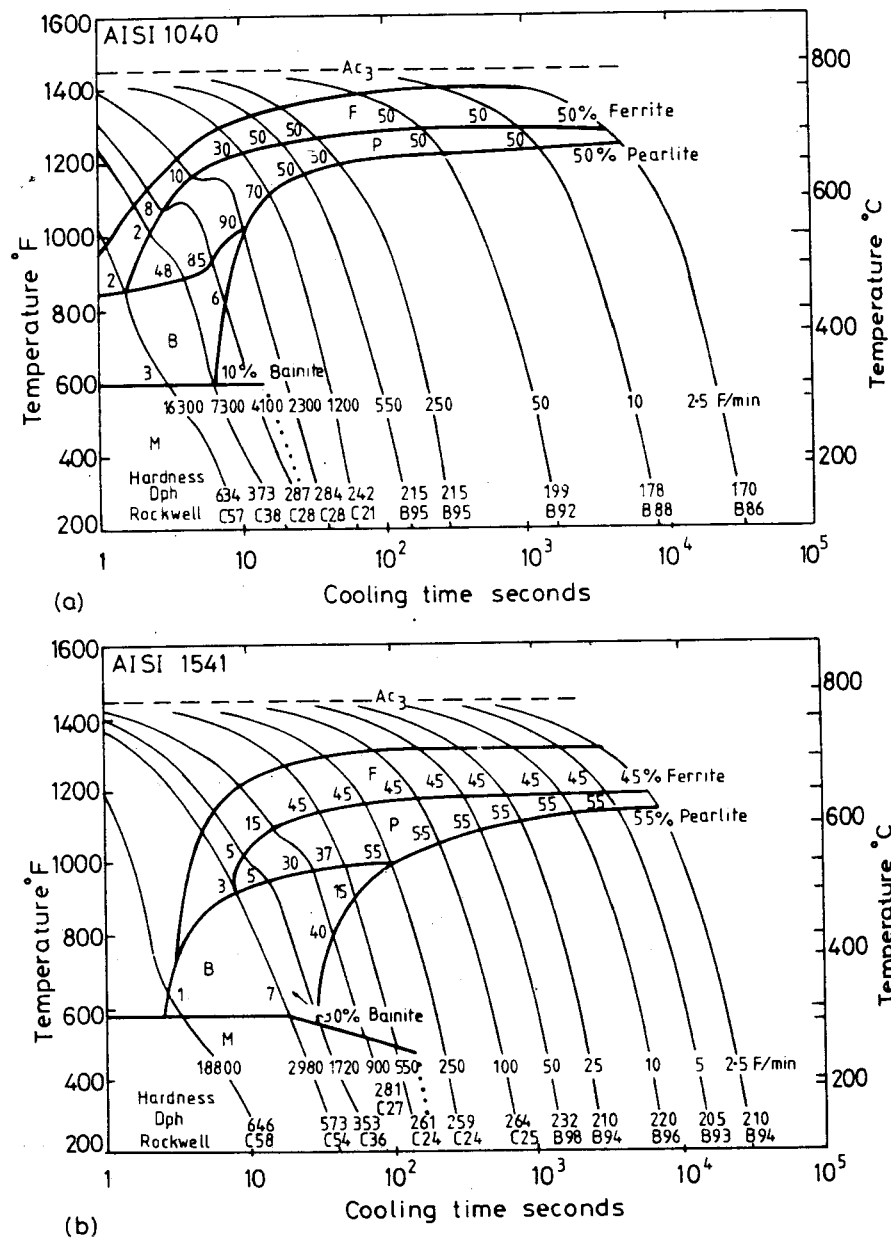


Fig. 5.70 CCT diagrams showing the influence of Mn on a 0.4 wt% C steel. (a) 0.39 C, 0.72 Mn, 0.23 Si, 0.018 S, 0.010 P. $Ac_1 = 728^\circ\text{C}$, $Ac_3 = 786^\circ\text{C}$. Grain size, ASTM No. 7–8. (b) 1.6 Mn: 0.39 C, 1.56 Mn, 0.21 Si, 0.024 S, 0.010 P. $Ac_1 = 716^\circ\text{C}$, $Ac_3 = 788^\circ\text{C}$. Grain size, ASTM No. 8. F, ferrite; P, pearlite; B, bainite; M, martensite. (Atlas of Isothermal Transformation and Cooling Transformation Diagrams, American Society of Metals, 1977, p. 414.)

5.8.5 Fibrous and Interphase Precipitation in Alloy Steels

When a few per cent of a strong carbide-forming element (e.g. Mo, W, Cr, Ti, V) is alloyed with steel, cementite is entirely replaced by a more stable carbide. When such steels are isothermally transformed at temperatures where the substitutional alloying element has appreciable mobility (~ 600 – 750°C) two new alloy-carbide morphologies can form.

Sometimes a fibrous morphology can be formed as illustrated in Fig. 5.71. This is a mixture of Mo_2C fibres in ferrite. The interfibre spacings are about an order of magnitude less than found in pearlite with fibre diameters ~ 10 – 50 nm.

In other cases planar arrays of alloy carbides in ferrite are produced, Fig. 5.72. The spacing of the sheets of precipitates decreases with decreasing temperature of transformation, being of the order of 10 – 50 nm. The sheets of precipitate are parallel to successive positions of the γ/α interface, hence this type of precipitation is known as *interphase precipitation*. The mechanism by which the microstructure develops is shown in the thin-foil electron micrographs in Fig. 5.73a and b and schematically in Fig. 5.73c. The α/γ interface can be seen to advance by the ledge mechanism, whereby mobile incoherent ledges migrate across immobile semicoherent facets. Note that these *growth* ledges are ~ 100 atom layers high in contrast to the *structural* ledges discussed in Section 3.4.1 which are only a few atom layers high at most. Normally the incoherent risers would be energetically favourable sites for precipitation, but in this case the ledges are moving too fast for nucleation to occur. As can be seen in Fig. 5.73a and as shown schematically in Fig. 5.73c, the precipitate size increases with distance behind the step, indicating that nucleation occurs on the semicoherent facets just ahead of the steps.

5.9 Massive Transformations

Consider the Cu–Zn alloys in Fig. 5.74 containing approximately 38 atomic % Zn. The most stable state for such alloys is β above $\sim 800^\circ\text{C}$, α below $\sim 500^\circ\text{C}$ and a mixture of $\alpha + \beta$ with compositions given by the equilibrium phase diagram in between. The type of transformation that occurs on cooling the β phase depends on the cooling rate. At slow to moderate cooling rates α precipitates in a similar way to the precipitation of ferrite from austenite in Fe–C alloys: slow cooling favours transformation at small undercooling and the formation of equiaxed α ; higher cooling rates result in transformation at lower temperatures and Widmanstätten α needles precipitate. According to the phase diagram, the α that precipitates will be richer in Cu than the parent β phase, and therefore the growth of the α phase requires the long-range diffusion of Zn away from the advancing α/β interfaces. This process is

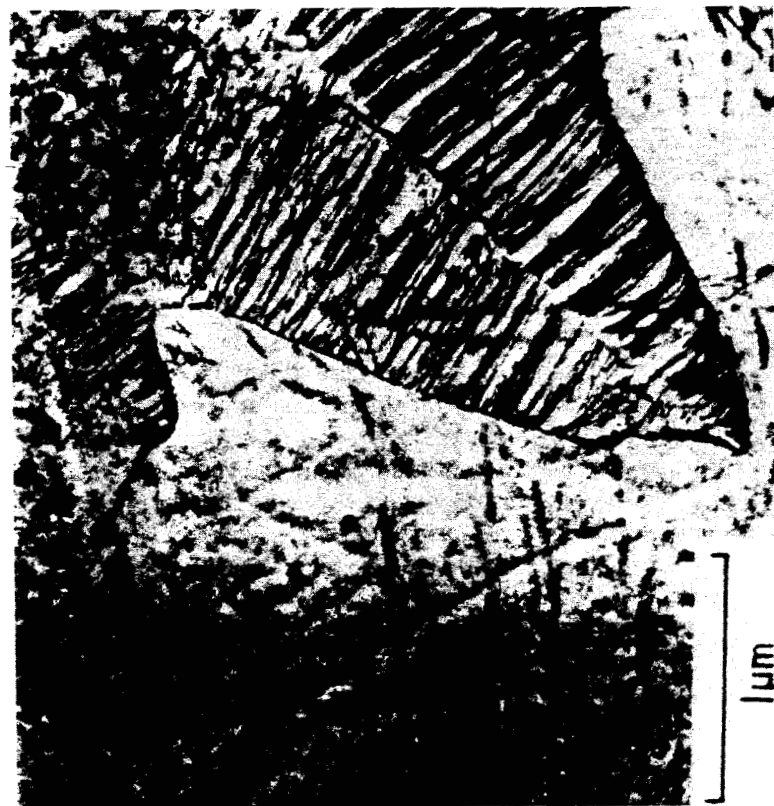


Fig. 5.71 Fe-4% Mo-0.2% C transformed 2 h at 650 °C. (a) Schematic of possible nucleation and growth mechanism. (b) Thin foil electron micrograph ($\times 32\,000$) (after R.W.K. Honeycombe, 'Transformation from Austenite in Alloy Steels', *Metallurgical Transactions*, 7A (1976) 91, © American Society for Metals and The Metallurgical Society of AIME, 1976, after D.V. Edmonds.)

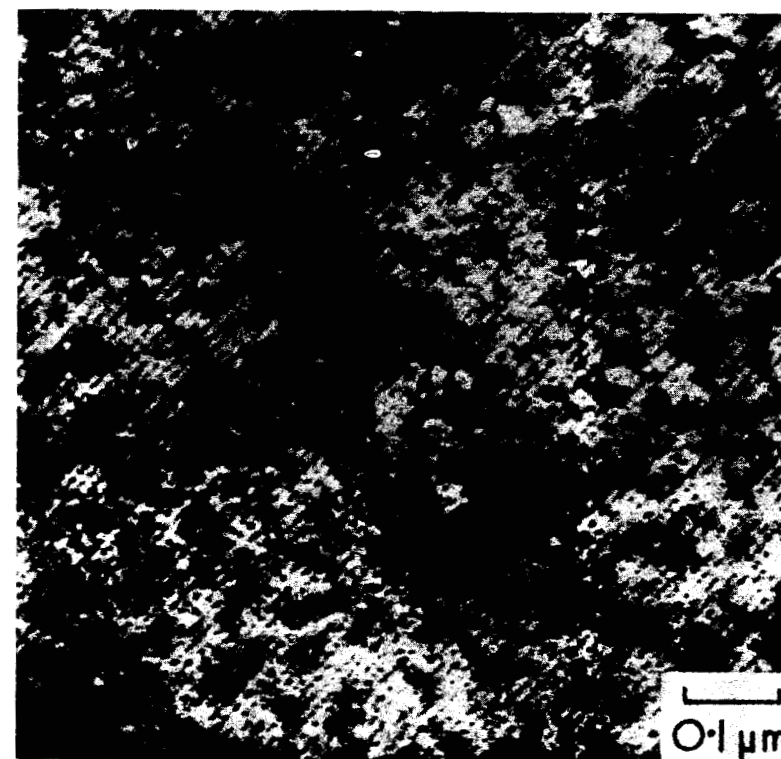
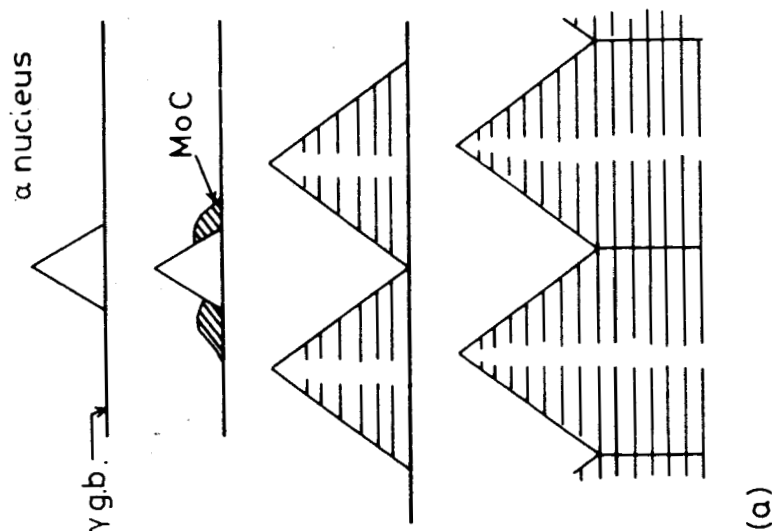
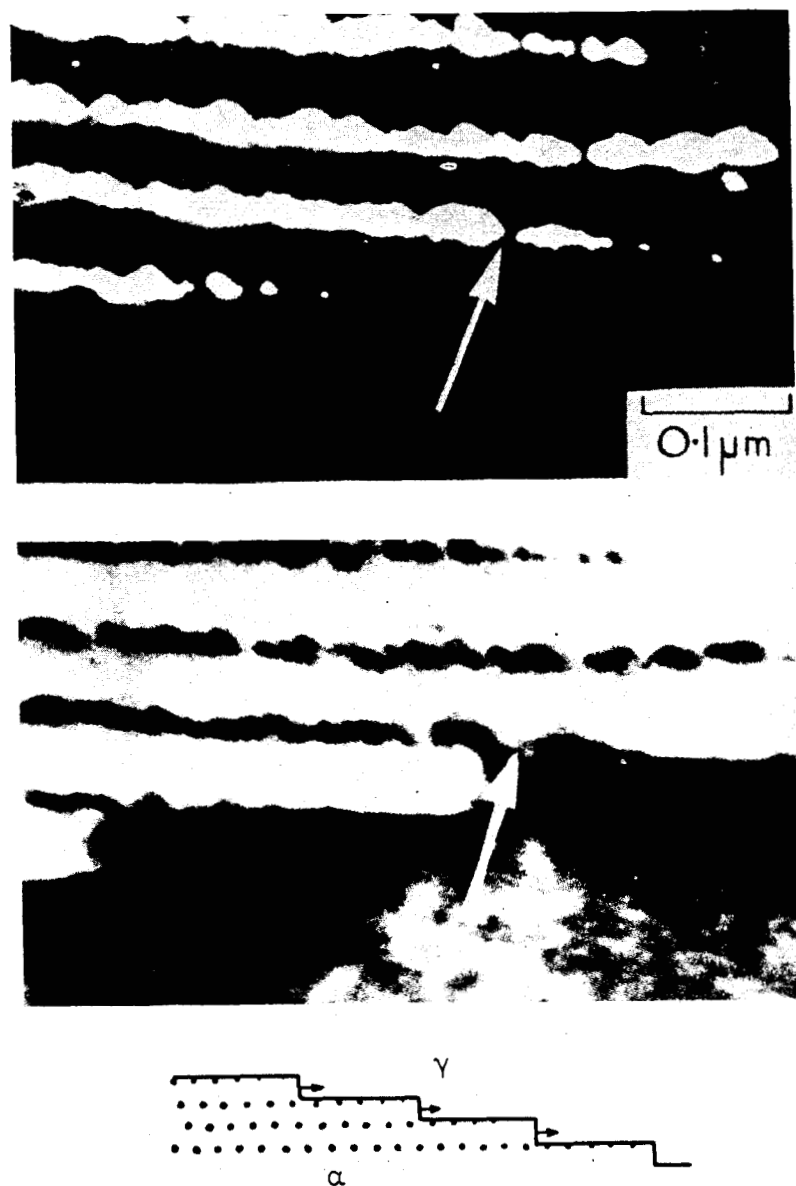


Fig. 5.72 Fe-0.75% V-0.15% C after 5 min at 725 °C. Thin foil electron micrograph showing sheets of vanadium carbide precipitates (interphase precipitation). (After R.W.K. Honeycombe, 'Transformation from Austenite in Alloy Steels', *Metallurgical Transactions*, 7A (1976) 91, © American Society for Metals and The Metallurgical Society of AIME, 1976, after A.D. Batte.)

relatively slow, especially since the Cu and Zn form substitutional solid solutions, and consequently the C curve for the α precipitation on a TTT or CCT diagram will be located at relatively long times. A possible CCT diagram is shown schematically in Fig. 5.75.

If the alloy is cooled fast enough, by quenching in brine for example, there is no time for the precipitation of α , and the β phase can be retained to temperatures below 500 °C where it is possible for β to transform into α with the same composition. The result of such a transformation is a new massive transformation product, Fig. 5.76.

Massive α grains nucleate at grain boundaries and grow rapidly into the surrounding β . Note also that because of the rapid growth the α/β boundaries have a characteristic irregular appearance. Since both the α and β phases have the same composition, massive α (α_m) can grow as fast as the Cu and Zn atoms can cross the α/β interface, without the need for long-range diffusion.



(c)

Fig. 5.73 Fe-12% Cr-0.2% C transformed 30 min at 650°C . Interphase precipitation of Cr_{23}C_6 at α/γ interface. (a) Dark-field micrograph showing bright precipitates. (b) Bright-field micrograph of same area showing ledges in the α/γ interface. Precipitates appear dark. (c) Schematic of nucleation and growth mechanism for interphase precipitation. (After R.W.K. Honeycombe, 'Transformation from Austenite in Alloy Steels', *Metallurgical Transactions* 7A (1976) 91. © American Society for Metals and The Metallurgical Society of AIME, 1976, after K. Campbell.)

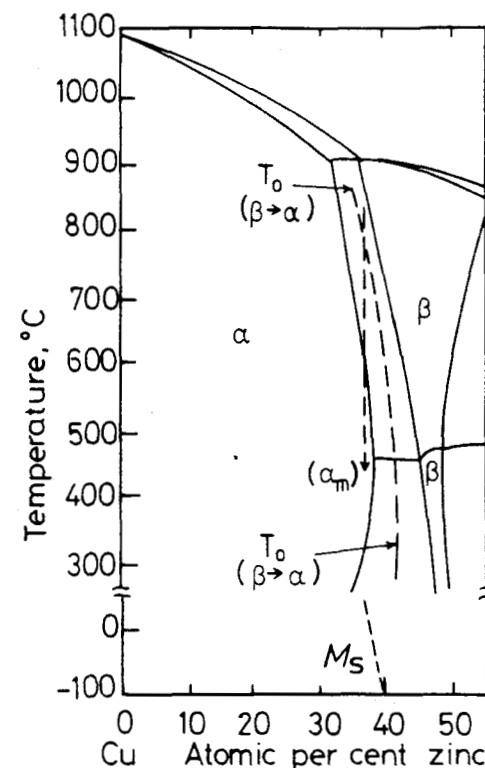


Fig. 5.74 A part of the Cu-Zn phase diagram showing the α/β equilibrium. The temperature at which $G^\alpha = G^\beta$ is marked as T_0 . M_s marks the beginning of the martensite transformation in rapidly quenched specimens. (After T.B. Massalski in *Phase Transformations*. American Society for Metals, 1970.)

Since growth only involves thermally activated jumping across the α/β interface, the massive transformation can be defined as a *diffusionless civilian transformation* and it is characterized by its own C curve on TTT or CCT diagrams as shown in Fig. 5.75. The migration of the α/β interfaces is very similar to the migration of grain boundaries during recrystallization of single-phase material. However, in the case of the massive transformation the driving force is orders of magnitude greater than for recrystallization, which explains why the transformation is so rapid.

Massive transformations should not be confused with martensite. Although the martensitic transformation also produces a change of crystal structure without a change in composition, the transformation mechanism is quite different. Martensite growth is a diffusionless *military* transformation, i.e. β is sheared into α by the cooperative movement of atoms across a glissile interface, whereas the growth of massive α involves thermally activated interface migration. Systems showing massive transformations will generally

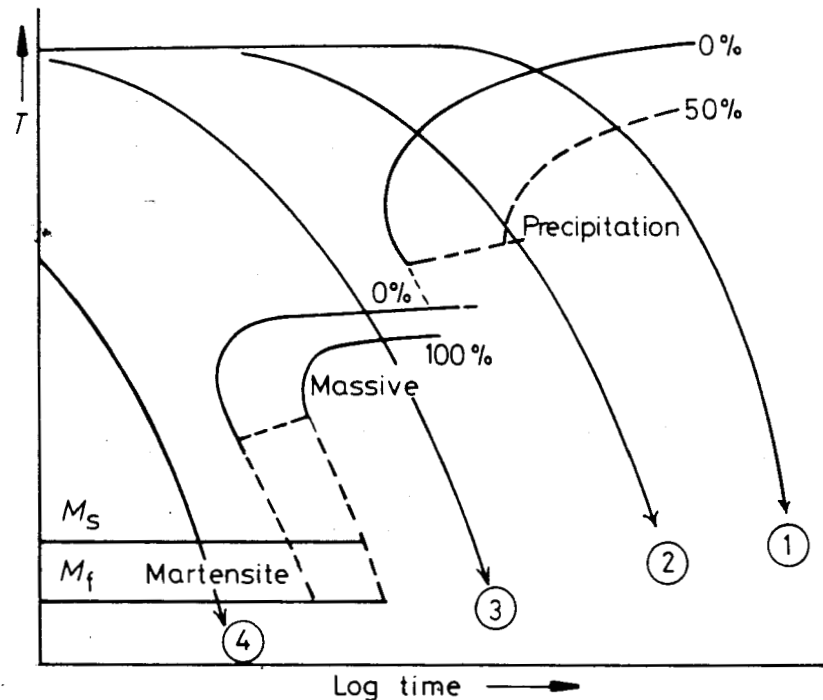


Fig. 5.75 A possible CCT diagram for systems showing a massive transformation. Slow cooling (1) produces equiaxed α . Widmanstätten morphologies result from faster cooling (2). Moderately rapid quenching (3) produces the massive transformation, while the highest quench rate (4) leads to a martensitic transformation. Compare with Fig. 5.79.

also transform martensitically if sufficiently high quench rates are used to suppress the nucleation of the massive product, Fig. 5.75. However Fig. 5.74 shows that for the Cu–Zn alloys the M_s temperature is below 0°C and some β phase is therefore retained after quenching to room temperature, as can be seen in Fig. 5.76.

It was stated above that β can transform massively into α provided the β phase could be cooled into the stable α phase field without precipitation at a higher temperature. Thermodynamically, however, it is possible for the transformation to occur at higher temperatures. The condition that must be satisfied for a massive transformation is that the free energy of the new phase must be lower than the parent phase, both phases having the same composition. In the case of Cu–38 atomic % Zn therefore, it can be seen from Fig. 5.77 that there is a temperature $\sim 700^\circ\text{C}$ below which G^α becomes less than G^β . This temperature is marked as T_0 in Fig. 5.74 and the locus of T_0 is also shown for other alloy compositions. Therefore it may be possible for a massive transformation to occur within the two-phase region of the phase diagram anywhere below the T_0 temperature. In practice, however, there is



Fig. 5.76 Massive α formed at the grain boundaries of β in Cu–38.7 wt% Zn quenched from 850°C in brine at 0°C . Some high temperature precipitation has also occurred on the boundaries. (From D. Hull and K. Garwood, *The Mechanism of Phase Transformations in Metals*, Institute of Metals, London, 1956.)

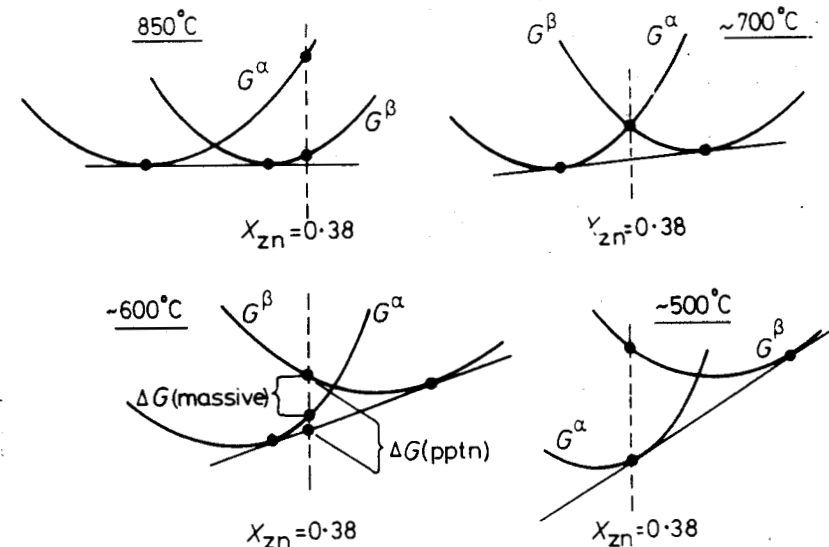


Fig. 5.77 A schematic representation of the free energy–composition curves for α and β in the Cu–Zn system at various temperatures.

evidence that massive transformations usually occur only within the single-phase region of the phase diagram. ?

Massive transformations are found in many alloy systems. Usually the interfaces are incoherent and migrate by continuous growth in a similar manner to a high-angle grain boundary, but in some cases growth can take place by the lateral movement of ledges across faceted interfaces. The Cu-Al phase diagram is similar to that shown in Fig. 5.74. Figure 5.78 shows a specimen of Cu-20 atomic % Al that has been quenched from the β field to produce almost 100% massive α . Again, characteristically irregular phase boundaries are apparent. In both Figs. 5.76 and 5.78 the cooling rate has been insufficient to prevent some precipitation on grain boundaries at higher temperatures before the start of the massive transformation.

The $\gamma \rightarrow \alpha$ transformation in iron and its alloys can also occur massively provided the γ is quenched sufficiently rapidly to avoid transformation near equilibrium, but slow enough to avoid the formation of martensite. The effect of cooling rate on the temperature at which transformation starts in pure iron is shown in Fig. 5.79. The microstructure of massive ferrite is shown in Fig. 5.80. Note the characteristically irregular grain boundaries.

Massive transformations are not restricted to systems with phase diagrams

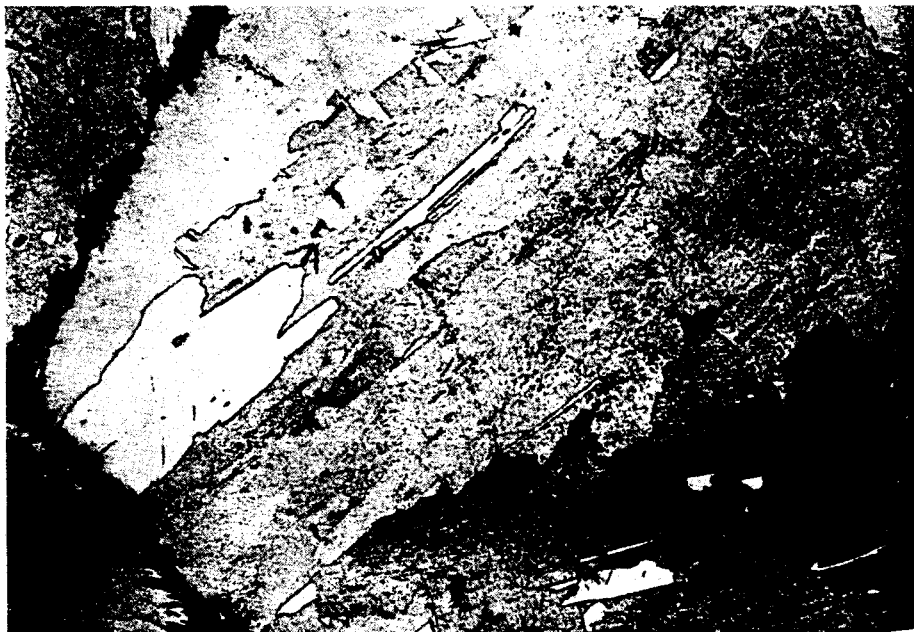


Fig. 5.78 Massive α in Cu-20 atomic % Al after quenching from the β field at 1027 °C into iced brine. Note the irregular α/α boundaries. Some other transformation (possibly bainitic) has occurred on the grain boundaries. (After G.A. Chadwick, *Metallography of Phase Transformations*, Butterworths, London, 1972.)

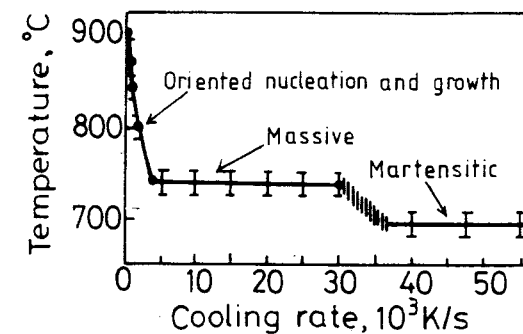


Fig. 5.79 The effect of cooling rate on the transformation temperature of pure iron. (After M.J. Bibby and J.G. Parr, *Journal of the Iron and Steel Institute* 202 (1964) 100.)

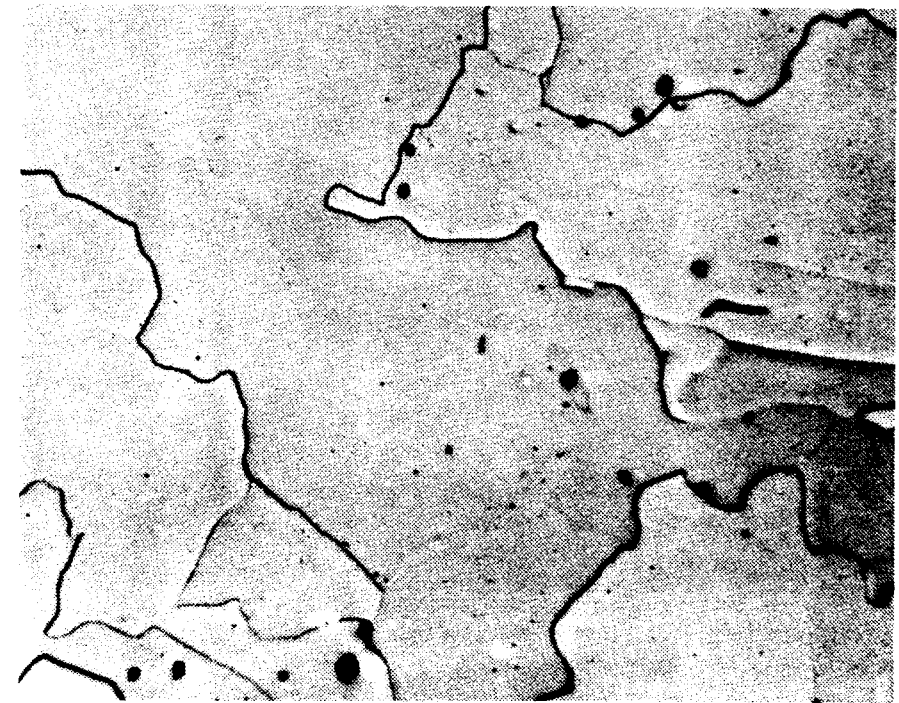


Fig. 5.80 Massive α in an Fe-0.002 wt% C quenched into iced brine from 1000 °C. Note the irregular α/α boundaries. (After T.B. Massalski in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 186.)

like that shown in Fig. 5.74. Metastable phases can also form massively as shown in Fig. 5.1d(ii) for example. It is not even necessary for the transformation product to be a single phase: two phases, at least one of which must be metastable, can form simultaneously provided they have the same composition as the parent phase.

5.10 Ordering Transformations

The structure of ordered phases has already been briefly discussed in Section 1.3.7. To recap: solid solutions which have a negative enthalpy of mixing ($\Omega < 0$) prefer unlike nearest neighbours and therefore show a tendency to form ordered phases at low temperatures. The five main types of ordered solutions are shown in Fig. 1.22. An example of a phase diagram containing low-temperature ordering reactions is the Au–Cu diagram in Fig. 1.21. Another example is the ordering of bcc β -brass below $\sim 460^\circ\text{C}$ to the so-called $L2_0$ (or $B2$) superlattice, Fig. 5.74. The bcc (or so-called $A2$) lattice can be considered as two interpenetrating simple cubic lattices: one containing the corners of the bcc unit cell and the other containing the body-centring sites. If these two sublattices are denoted as A and B the formation of a perfectly ordered β' superlattice involves segregation of all Cu atoms to the A sublattice, say, and Zn to the B sublattice. This is not feasible in practice, however, as the β' does not have the ideal CuZn composition. There are two ways of forming ordered structures in non-stoichiometric phases: either some atom sites can be left vacant or some atoms can be located on wrong sites. In the case of $\beta(\text{CuZn})$ the excess Cu atoms are located on some of the Zn sites.

Let us begin the discussion of ordering transformations by considering what happens when a completely ordered single crystal such as CuZn or Cu_3Au is heated from low temperatures to above the disordering temperature. To do this it is useful to quantify the degree of order in the crystal by defining a long-range order parameter L such that $L = 1$ for a fully ordered alloy where all atoms occupy their 'correct' sites and $L = 0$ for a completely random distribution. A suitable definition of L is given by

$$L = \frac{r_A - X_A}{1 - X_A} \quad \text{or} \quad \frac{r_B - X_B}{1 - X_B}$$

where X_A is the mole fraction of A in the alloy and r_A is the probability that an A sublattice site is occupied by the 'right' kind of atom.

At absolute zero the crystal will minimize its free energy by choosing the most highly ordered arrangement ($L = 1$) which corresponds to the lowest internal energy. The configurational entropy of such an arrangement, however, is zero and at higher temperatures the minimum free energy state will contain some disorder, i.e. some atoms will interchange positions by diffusion so that they are located on 'wrong' sites. Entropy effects become increasingly more important with rising temperature so that L continuously decreases until above some critical temperature (T_c) $L = 0$. By choosing a suitable model, such as the quasi-chemical model discussed in Section 1.3.4, it is possible to calculate how L varies with temperature for different superlattices. The results of such a calculation for the CuZn and Cu_3Au superlattices are shown in Fig. 5.81. It can be seen that the way in which L decreases to

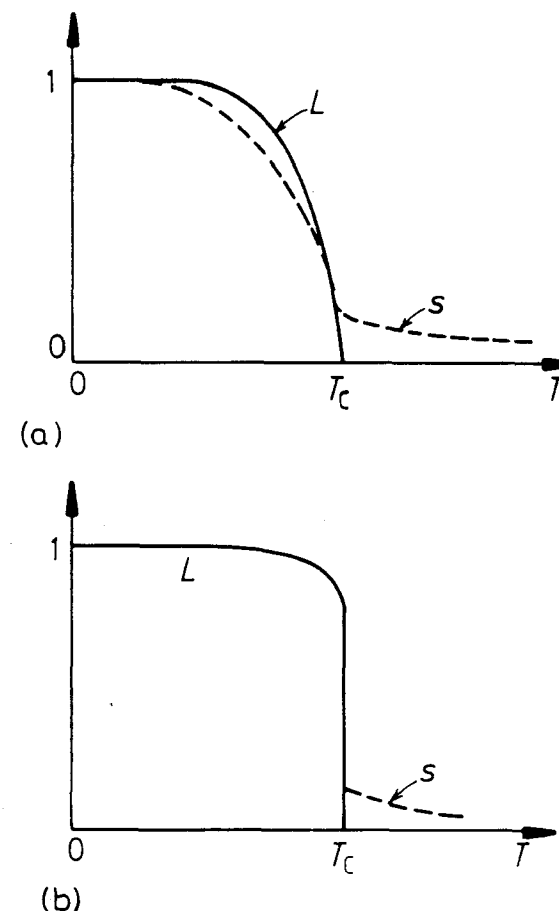


Fig. 5.81 The variation of long-range order (L) and short-range order (s) for (a) CuZn-type and (b) Cu_3Au -type transformations (schematic).

zero is different for the different superlattices. In the equiatomic CuZn case L decreases continuously with temperature up to T_c , whereas in Cu_3Au L decreases only slightly up to T_c and then abruptly drops to zero above T_c . This difference in behaviour is a consequence of the different atomic configurations in the two superlattices.

Above T_c it is impossible to distinguish separate sublattices extending over long distances and $L = 0$. However, since $\Omega < 0$ there is still a tendency for atoms to attract unlike atoms as nearest neighbours, i.e. there is a tendency for atoms to order over short distances. The degree of short-range order (s) is defined in Section 1.3.7. The variation of s with temperature is shown as the dashed lines in Fig. 5.81.

The majority of phase transformations that have been discussed in this book have been so-called *first-order transformations*. This means that at the

equilibrium transformation temperature the first derivatives of the Gibbs free energy $\partial G/\partial T$ and $\partial G/\partial P$ are discontinuous. The melting of a solid is such a transformation. Fig. 5.82a. Since $\partial G/\partial T = -S$ and $\partial G/\partial P = V$, first order transformations are characterized by discontinuous changes in S and V . There is also a discontinuous change in enthalpy H corresponding to the evolution of a latent heat of transformation. The specific heat of the system is effectively infinite at the transformation temperature because the addition of a small

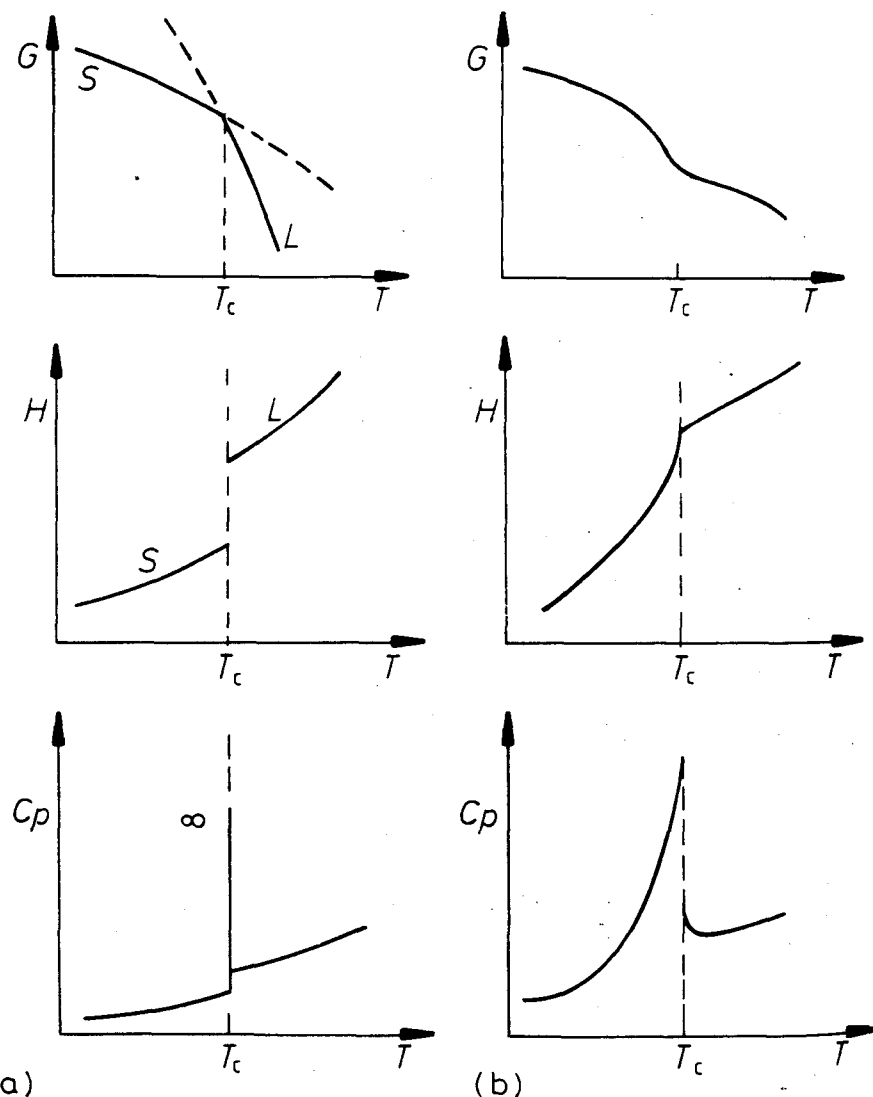


Fig. 5.82 The thermodynamic characteristics of (a) first-order and (b) second-order phase transformations.

quantity of heat converts more solid into liquid without raising the temperature.

Figure 5.82b illustrates the characteristics of a *second-order transformation*. For such a transformation the *second derivatives of Gibbs free energy* $\partial^2 G/\partial T^2$ and $\partial^2 G/\partial P^2$ are discontinuous. The first derivatives, however, are continuous which means that H is also continuous. Consequently since

$$\left(\frac{\partial^2 G}{\partial T^2}\right)_P = -\left(\frac{\partial S}{\partial T}\right)_P = \frac{1}{T} \left(\frac{\partial H}{\partial T}\right)_P = \frac{C_P}{T}$$

there is no latent heat, only a high specific heat, associated with the transformation.

Returning to a consideration of order-disorder transformations it can be seen from Fig. 5.81 that the loss of long-range order in the $\beta' \rightarrow \beta$ (CuZn) transformation corresponds to a gradual disordering of the structure over a range of temperatures. There is no sudden change in order at T_c and consequently the internal energy and enthalpy (H) will be continuous across T_c . The $\beta' \rightarrow \beta$ transformation is therefore a second-order transformation. In the case of Cu_3Au , on the other hand, a substantial change in order takes place discontinuously at T_c . Since the disordered state will have a higher internal energy (and enthalpy) than the ordered state, on account of the greater number of high-energy like-like atom bonds, there will be a discontinuous change in H at T_c , i.e. the transformation is first order.

So far we have been concerned with the disordering transformation that takes place on heating a fully ordered single crystal. The mechanism by which order is lost is most likely the interchange of atoms by diffusional processes occurring homogeneously throughout the crystal. The same changes will of course take place in every grain of a polycrystal. Let us now turn to the reverse transformation that occurs on cooling a single crystal, i.e. disorder $\rightarrow$ order.

There are two possible mechanisms for creating an ordered superlattice from a disordered solution. (1) There can be a continuous increase in short-range order by local rearrangements occurring homogeneously throughout the crystal which finally leads to long-range order. (2) There may be an energy barrier to the formation of ordered domains, in which case the transformation must take place by a process of *nucleation and growth*. These two alternative mechanisms are equivalent to spinodal decomposition and precipitation as mechanisms for the formation of coherent zones in alloys with positive heats of mixing ($\Omega > 0$). The first mechanism may only be able to operate in second-order transformations or at very high supercoolings below T_c . The second mechanism is generally believed to be more common.

The nucleation and growth process is illustrated in Fig. 5.83. The disordered lattice is represented by the cross-grid of lines. Within this lattice two sublattices are marked by heavy and faint lines. Atoms are located at each intersection but only atoms within the ordered regions, or domains, are

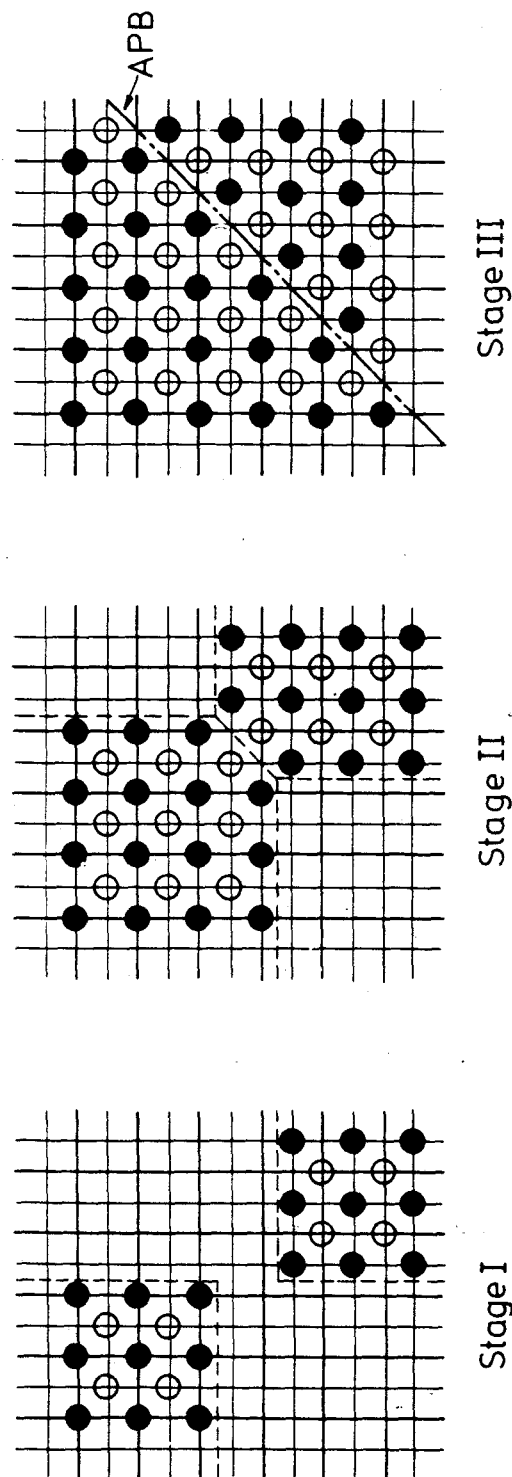


Fig. 5.83 The formation of an antiphase boundary (APB) when out-of-phase ordered domains grow together. The diagram could represent a {100} plane in Cu_3Au in which case the black and white atoms could represent Cu and Au. (After P. Haasen, *Physical Metallurgy*, Cambridge University Press, Cambridge, 1978.)

marked; the unmarked sites are disordered. The diagram is only schematic, but could represent a {100} plane of the Cu_3Au superlattice. Since the two types of atoms can order on either the A or B sublattice, the independently nucleated domains will often be 'out of phase' as shown. When these domains subsequently grow together a boundary will form (known as an antiphase domain boundary or APB) across which the atoms will have the wrong kind of neighbours. APBs are therefore high-energy regions of the lattice and are associated with an APB energy.

Even at rather low undercoolings below T_c the activation energy barrier to the nucleation of ordered domains ΔG^* should be rather small because both nucleus and matrix have essentially the same crystal structure and are therefore coherent with a low interfacial energy. Also, provided the alloy has a stoichiometric composition, both nucleus and matrix have the same composition so that there should not be large strain energies to be overcome. Consequently, it is to be expected that nucleation will be homogeneous, independent of lattice defects such as dislocations and grain boundaries. Figure 5.84 shows evidence for the existence of a nucleation and growth mechanism during ordering in CoPt. This is a field ion micrograph showing that the two types of atoms are ordered in a regular manner in the upper part but disordered in the lower part of the micrograph.

At low ΔT the nucleation rate will be low and a large mean domain size results, whereas higher values of ΔT should increase the nucleation rate and diminish the initial domain size. The degree of long-range order in a given domain will vary with temperature according to Fig. 5.81 and with decreasing temperature the degree of order is increased by homogeneous diffusive rearrangements among the atoms within the domain. Within the crystal as a whole, the degree of long-range order will initially be very small because there are likely to be equal numbers of domains ordered on both A and B lattices. The only way for long-range order to be established throughout the entire crystal is by the coarsening of the APB structure. The rate at which this occurs depends on the type of superlattice.

In the CuZn -type superlattice (L_{20}) there are only two sublattices on which the Cu atoms, say, can order and therefore only two distinct types of ordered domain are possible. A consequence of this is that it is impossible for a metastable APB structure to form. It is therefore relatively easy for the APB structure to coarsen in this type of ordered alloy. Figure 5.85 shows an electron micrograph of APBs in AlFe (L_{20} superlattice) along with a schematic diagram to illustrate the two different types of domain. The Cu_3Au (L_{12}) superlattice is different to the above in that there are four different ways in which ordered domains can be formed from the disordered fcc lattice: the Au atoms can be located either at the corners of the unit cell, Fig. 1.20c, or at the one of the three distinct face-centred sites. The Cu_3Au APBs are therefore more complex than the CuZn type, and a consequence of this is that it is possible for the APBs to develop a metastable, so-called foam structure, Fig. 5.86. Another interesting feature of this microstructure is that the APBs

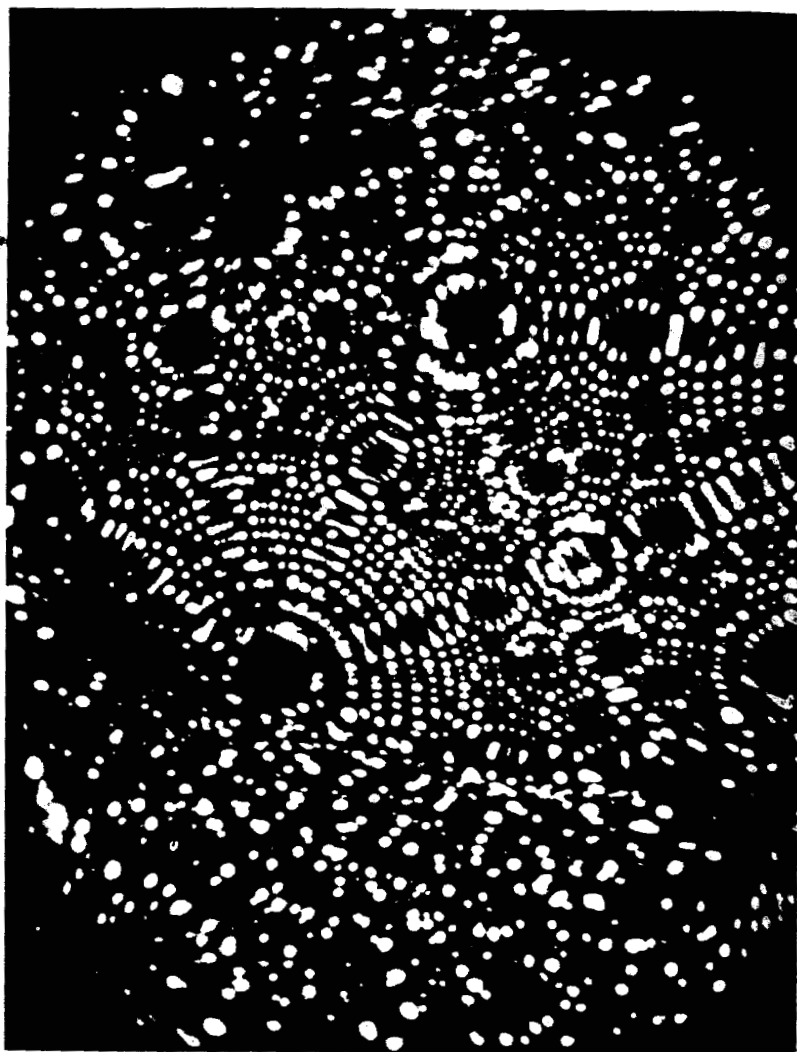


Fig. 5.84 A field ion micrograph of the boundary between an ordered domain (above) and disordered matrix (below) in CoPt. (After E.W. Müller and Tien Tzong Tsong, *Field Ion Microscopy, Principles and Applications*, Elsevier, 1969.)

tend to align parallel to $\{100\}$ planes in order to minimize the number of high-energy Au-Au bonds.

The rate at which ordering occurs varies greatly from one alloy to another. For example the ordering of $\beta(\text{CuZn})$ is so rapid that it is almost impossible to quench-in the disordered bcc structure. This is because the transformation is second order and can occur by a rapid continuous ordering process.

Ordering of Cu_3Au on the other hand is relatively slow requiring several hours for completion, despite the fact that the atomic mobilities ought to be

similar to those in the CuZn transformation. This transformation, however, is second order and proceeds by nucleation and growth. Also the development of long-range order is impeded by the formation of metastable APB networks.

The above comments have been concerned primarily with alloys of stoichiometric composition. However, it has already been pointed out that ordering is often associated with non-stoichiometric alloys.

In the case of first-order transformations there is always a two-phase region at non-stoichiometric compositions, Fig. 1.21, so that the transformation can be expressed as: disordered phase $\rightarrow$ ordered precipitates + disordered matrix. There is then a change in composition on ordering and long-range diffusion must be involved. Second-order transformations on the other hand do not involve a two-phase region even at non-stoichiometric compositions, Fig. 5.74.



Fig. 5.85 (a) A thin-foil electron micrograph showing APBs in an ordered AlFe alloy ($\times 17\,000$). (b) A schematic representation of the atomic configurations comprising the APB structure in (a). (After M.J. Marcinkowski in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 205.)

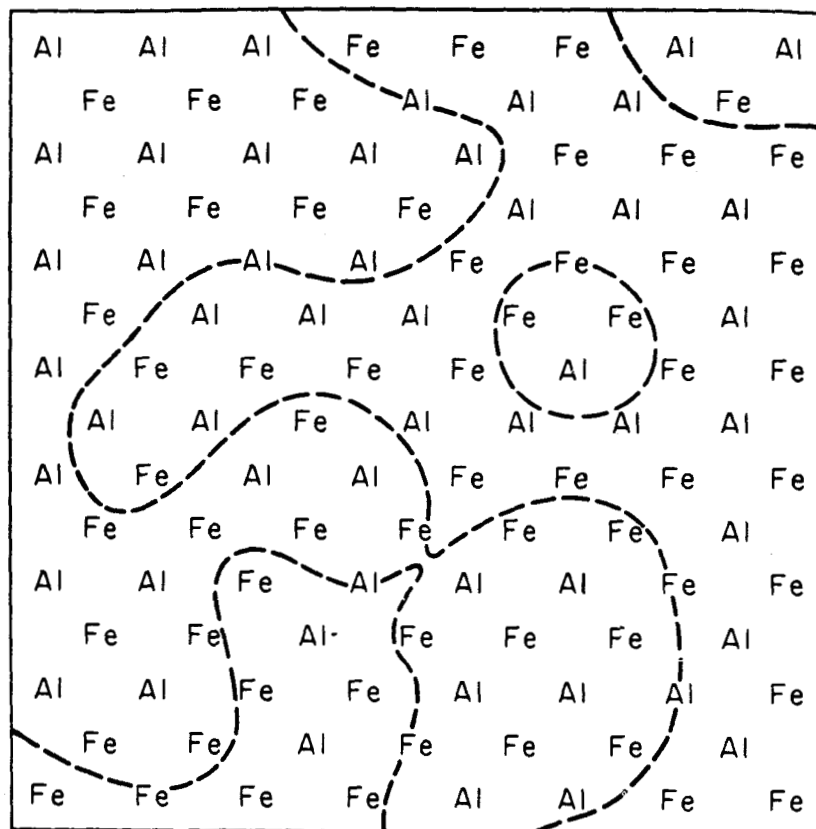


Fig. 5.85 (b)

5.11 Case Studies

5.11.1 Titanium Forging Alloys

Composition: Ti-6 wt% Al-4 wt% V.

Phase diagrams: Binary Ti-Al and Ti-V diagrams in Fig. 5.87.

Important phases: α -hcp, β -bcc.

Microstructures: See Figs. 5.88-5.91.

Applications: As a result of the high cost of titanium, uses are restricted to applications where high performance is required and high strength to weight ratio is important, e.g. gas turbine aero engines and airframe structures.

Comments: At low temperatures pure titanium exists as the hcp α phase, but above 883 °C up to the melting point (~1672 °C) the bcc β phase is stable. Figure 5.87 shows that Al is an α stabilizer, i.e. it raises the α/β transition temperature, whereas V is a β stabilizer which lowers the transition tempera-

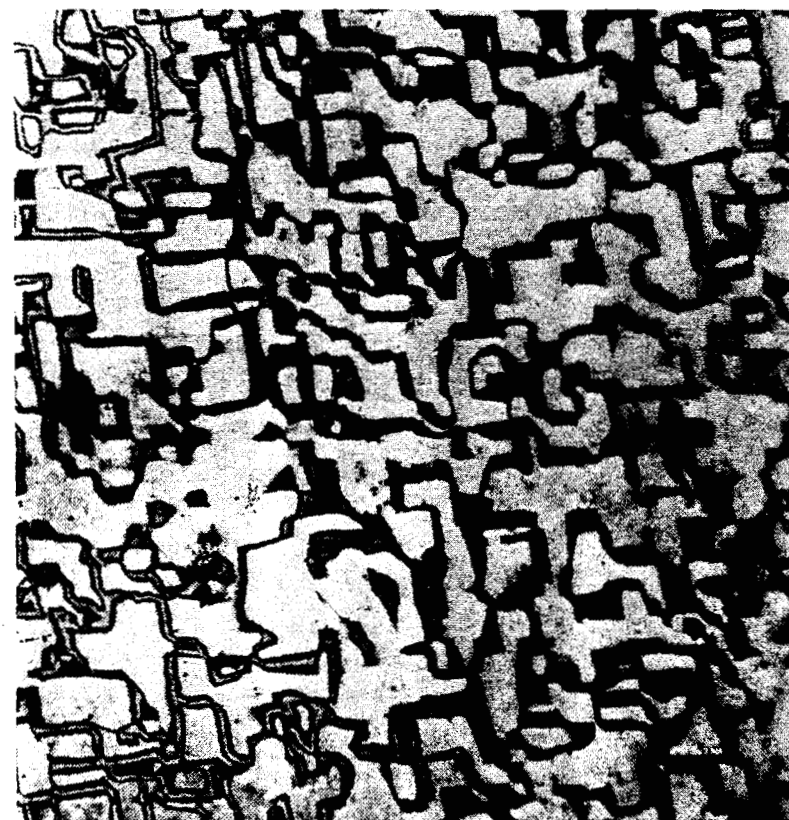


Fig. 5.86 APBs in ordered Cu_3Au . Thin-foil electron micrograph $\times 53\,000$. Note that due to the method of imaging about one third of the APBs are invisible. (After M.J. Marcinkowski in *Metals Handbook*, 8th edn., Vol. 8. American Society for Metals, 1973, p. 205.)

ture. A wide range of titanium alloys are available. These can be classified as either α , $\alpha + \beta$, or β alloys. The Ti-6Al-4V alloy to be discussed here belongs to the $\alpha + \beta$ group of alloys. For simplicity, the phase diagram relevant to these alloys can be envisaged as that shown in Fig. 5.87c. Two principal types of transformation are of interest. The first of these is the precipitation of α from β on cooling from above the β transus into the $\alpha + \beta$ field. This is in principle the same as the formation of ferrite during the cooling of austenite in Fe-C alloys. However, in this case the Widmanstätten morphology predominates at all practical cooling rates, Fig. 5.88a. Figure 5.88b is a thin-foil electron micrograph of a similar structure and shows more clearly the two phases present after air cooling. The β phase remains as a thin layer between the Widmanstätten α plates. Furnace cooling produces similar though coarser microstructures. The α plates and the β

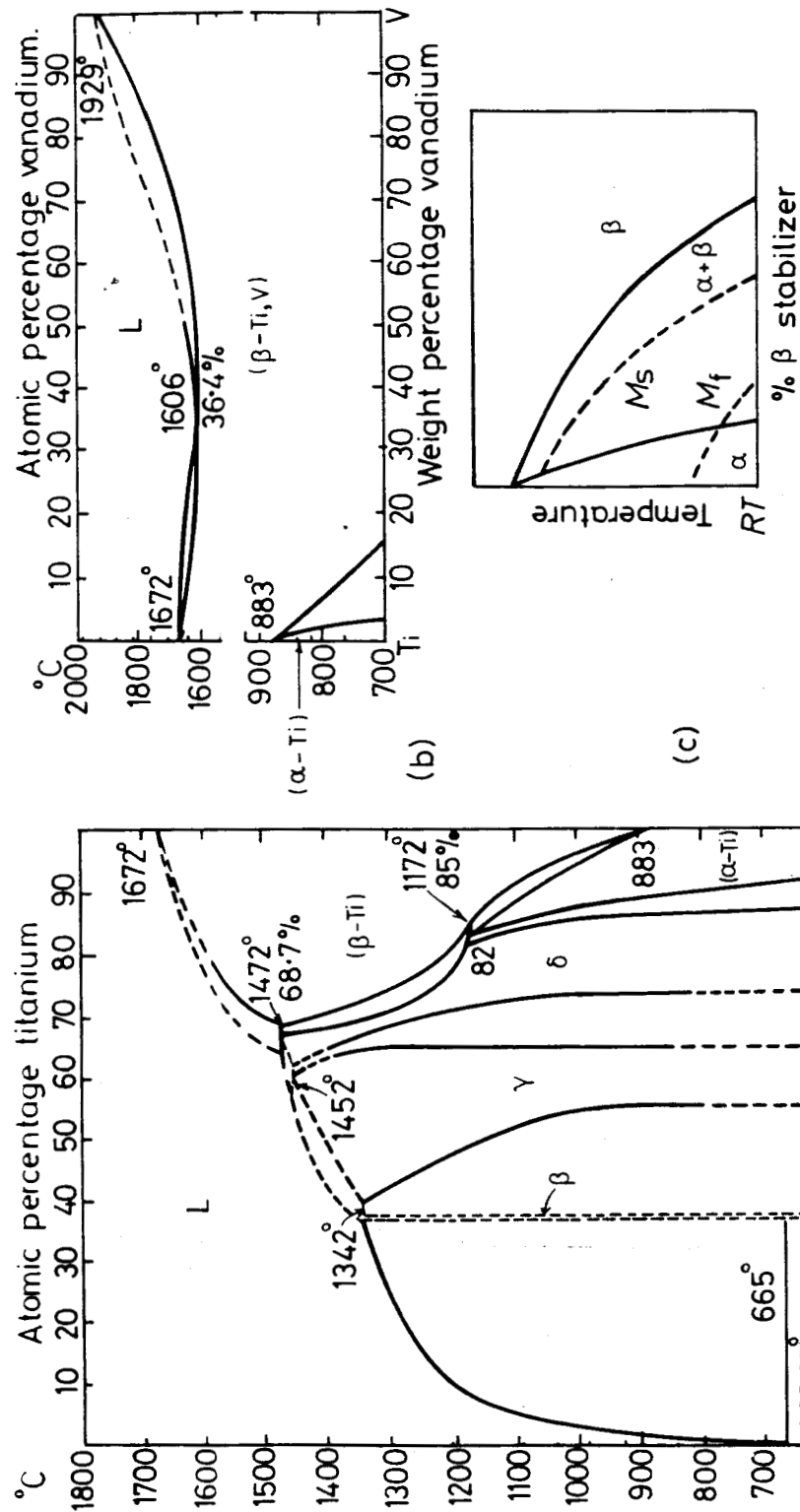


Fig. 5.87 Phase diagrams: (a) Ti-Al, (b) Ti-V, (c) a schematic diagram showing the variation of M_s and M_f with percentage β stabilizer. [(a) and (b) After H. Margelin in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 264.]

matrix are oriented such that

$$\begin{aligned} (0001)_\alpha // (110)_\beta \\ [11\bar{2}0]_\alpha // [\bar{1}11]_\beta \end{aligned}$$

The second important transformation is the formation of martensite that takes place when β is rapidly cooled by water quenching. The transformation can be written $\beta \rightarrow \alpha'$ where α' is a supersaturated hcp α phase. M_f for β containing 6 wt% Al-4 wt% V is above room temperature so that quenching from above the β transus produces a fully martensitic structure, Fig. 5.89a. The martensite can be aged by heating to temperatures where appreciable diffusion can occur, in which case the supersaturated α' can decompose by the precipitation of β on the martensite plate boundaries and dislocations, Fig. 5.89b.

Alloys for engineering applications are not usually used in the above conditions but are hot worked in the α + β region of the phase diagram in

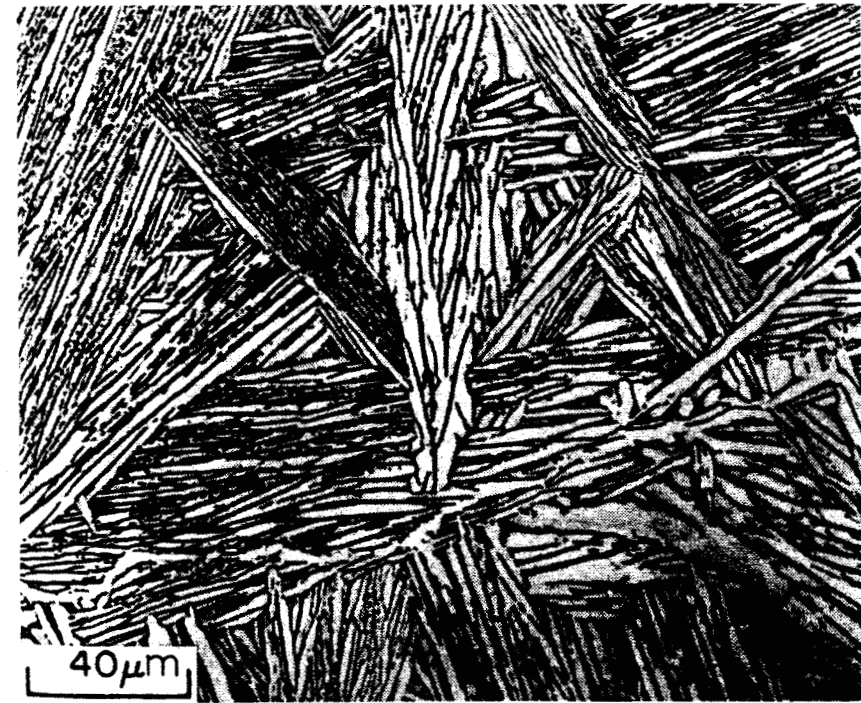


Fig. 5.88 (a) Widmanstätten α (light) and β (dark) in a Ti-6 Al-4 V alloy air cooled from 1037 °C. (S.M. Copley and J.C. Williams, in *Alloy and Microstructural Design* J.K. Tien and G.S. Ansell (Eds.), Academic Press, 1976.) (b) Alternate layers of α (light) and β (dark) in a Widmanstätten microstructure. Ti-6 Al-4 V forged at 1038 °C above the β transus, air-cooled, annealed 2 h at 704 °C, air cooled. Thin-foil electron micrograph (× 15 000). (From *Metals Handbook*, 8th edn., Vol. 7, American Society for Metals, 1972, p. 00.)

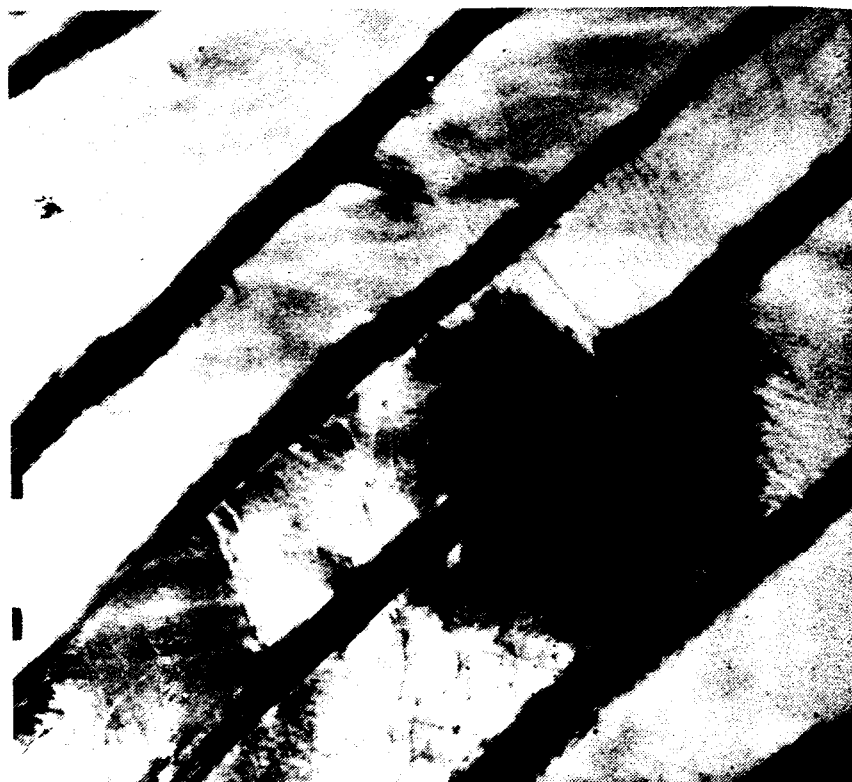


Fig. 5.88 (b)

order to break up the structure and distribute the α phase in a finely divided form. This is usually followed by annealing at 700 °C which produces a structure of mainly α with finely distributed retained β , Fig. 5.90. The advantage of this structure is that it is more ductile than when the α is present in a Widmanstätten form. When additional strength is required the alloys are hardened by heating to high temperatures in the $\alpha + \beta$ range (~940 °C) so that a large volume fraction of β is produced, followed by a water quench to convert the β into α' martensite, and then heating to obtain precipitation hardening of the martensite (Fig. 5.91). Mechanical properties that can be obtained after these treatments are given in Table 5.4. If the alloy is held

Table 5.4 Room Temperature Mechanical Properties of Ti—6 wt% Al—4 wt% V Alloys

Condition	YS/MPa	UTS/MPa	Elongation
annealed	930	990	15%
hardened	950	1030	14%

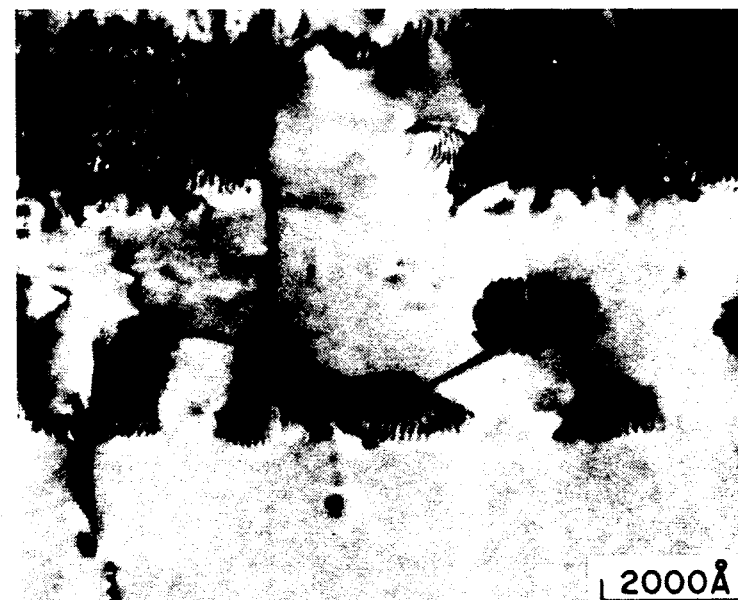
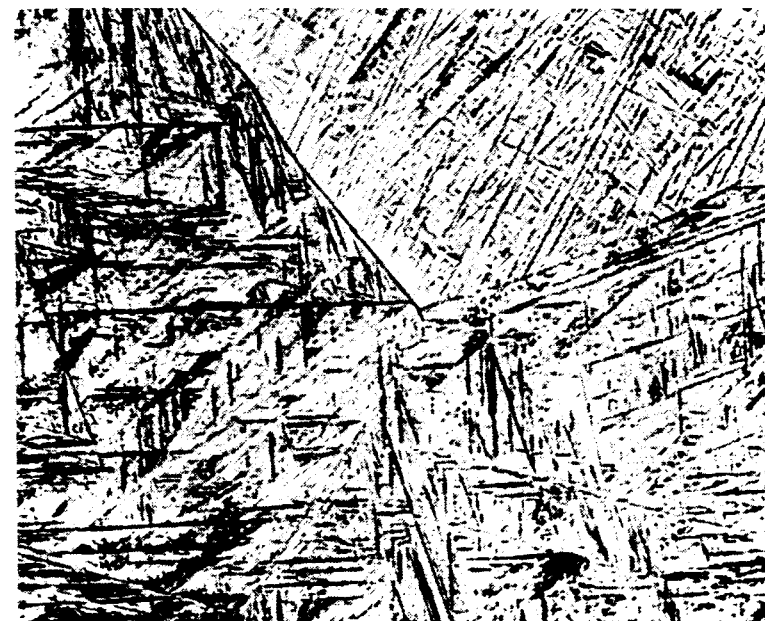


Fig. 5.89 (a) α' Martensite in Ti-6 Al-4 V held above the β transus at 1066 °C and water quenched. Prior β grain boundaries are visible ($\times 370$). (From *Metals Handbook*, 8th edn., Vol. 7, American Society for Metals, 1972, p. 328) (b) β precipitates that have formed during the tempering of α' martensite in Ti-6 Al-4 V. Specimen quenched from 1100 °C and aged 24 h at 600 °C. Thin-foiled electron micrograph. (After S.M. Copley and J.C. Williams in *Alloy and Microstructural Design*, J.K. Tien and G.S. Ansell (Eds.), Academic Press, New York, 1976.)



Fig. 5.90 Microstructure of hot-worked and annealed Ti-6 Al-4 V ($\times 540$). (P.H. Morton in *Rosenhain Centenary Conference*, The Royal Society, London, 1976.)

lower in the $\alpha + \beta$ field before quenching the β phase that forms can be so rich in vanadium that the M_s temperature is depressed to below room temperature and quenching results in retained β , see Fig. 5.87c.

5.11.2 The Weldability of Low-Carbon and Microalloyed Rolled Steels

Composition: C ≤ 0.22 wt%, Si $\approx 0.3\%$, Mn = 1.0–1.5%, P $\leq 0.04\%$, S $\leq 0.04\%$. C_{eq} (see text) $\leq 0.4\%$.

Possible microalloying elements: Al, Nb, Ti, V, with possible additions of Zr and/or N. The total amount of microalloying elements does not usually exceed 0.15%.

Phase diagrams: Fe–C binary.

Modified CCT diagrams (see below).

Welding nomographs (see text).

Microstructure: Depends on type of steel, e.g. whether quench and tempered, microalloyed—fine grained, plain rolled C–Mn, etc. See, e.g. Fig. 5.49a.

Applications: Constructional steels for building frames, bridges, pressure vessels, ships, oil platforms, etc.

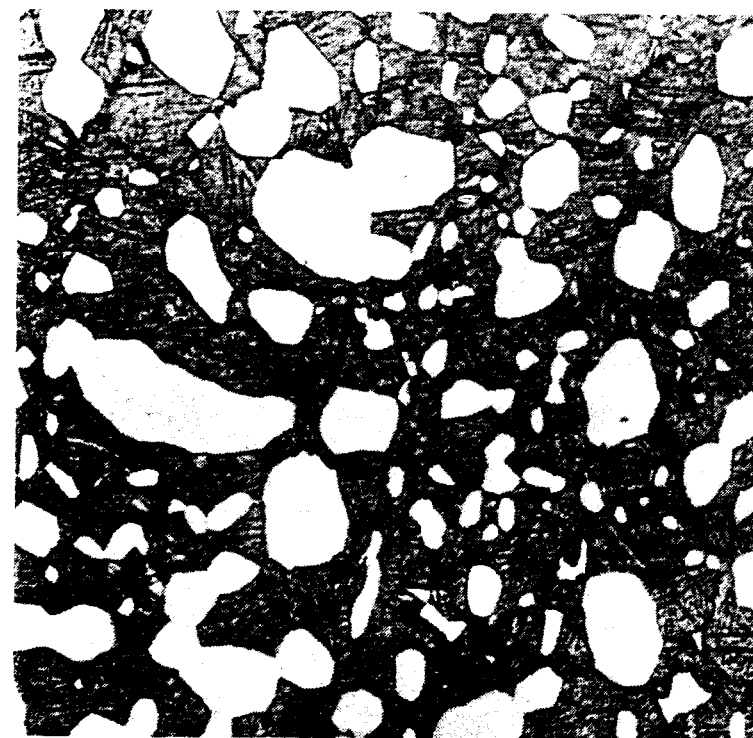


Fig. 5.91 Microstructure of hardened Ti-6 Al-4 V. Solution treated at 954 °C (high in the $\alpha - \beta$ range), water quenched, aged 4 h at 538 °C. Equiaxed 'primary' α grains (light) in an aged martensitic matrix. (From *Metals Handbook*, 8th edn., Vol. 7, American Society for Metals, 1972, p. 329.)

Comments: Steels used for heavy, high-strength constructions are nowadays rather sophisticated, relying for their high strength and toughness on having a fine and uniform grain size. When fusion welding plates together, the steel is subjected to an extremely severe thermal cycle, and at the fusion line the temperature attains the melting point of the alloy. Because the steel plate provides an effective thermal sink (see Section 4.5) the cooling rate is very high for most types of welding process as illustrated in Fig. 5.92. This thermal cycle causes changes in properties of the base material in the heat-affected zone due to the combination of phase changes and thermal/mechanical stresses. Typical microstructural changes experienced by a C–Mn steel are illustrated in Fig. 5.93, showing that recrystallization, grain growth and even ageing are occurring in the heat-affected zone. Of these changes, grain growth is potentially the most troublesome in decreasing the strength and toughness of these steels particularly since in most cases high-energy submerged arc welding is used with its associated relatively long dwell-time at peak temperatures. In order to avert the problem of grain growth at high temperatures, new steels have recently been introduced containing a fine dispersion of TiN

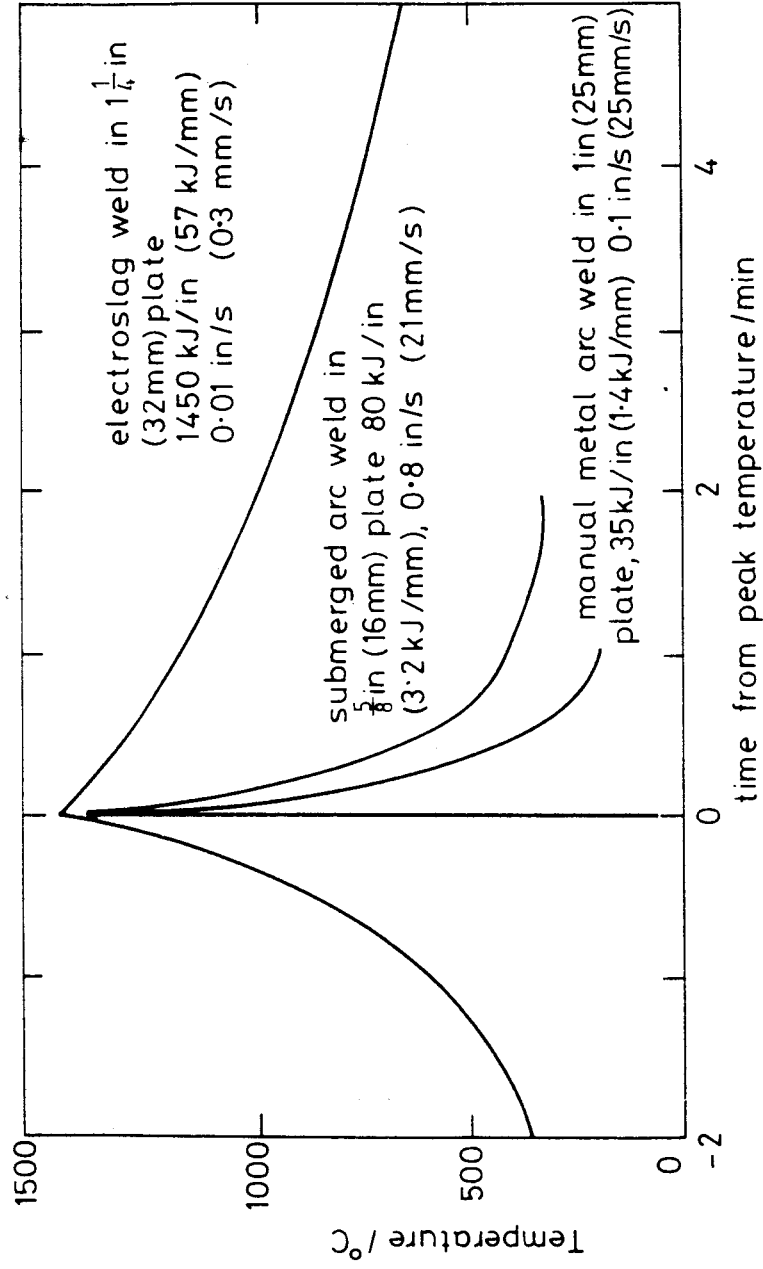


Fig. 5.92 Comparison of thermal cycles experienced in the heat-affected zone when welding steel by different processes. (After R.G. Baker in *Rosenhain Centenary Conference*, The Royal Society, London, 1976.)

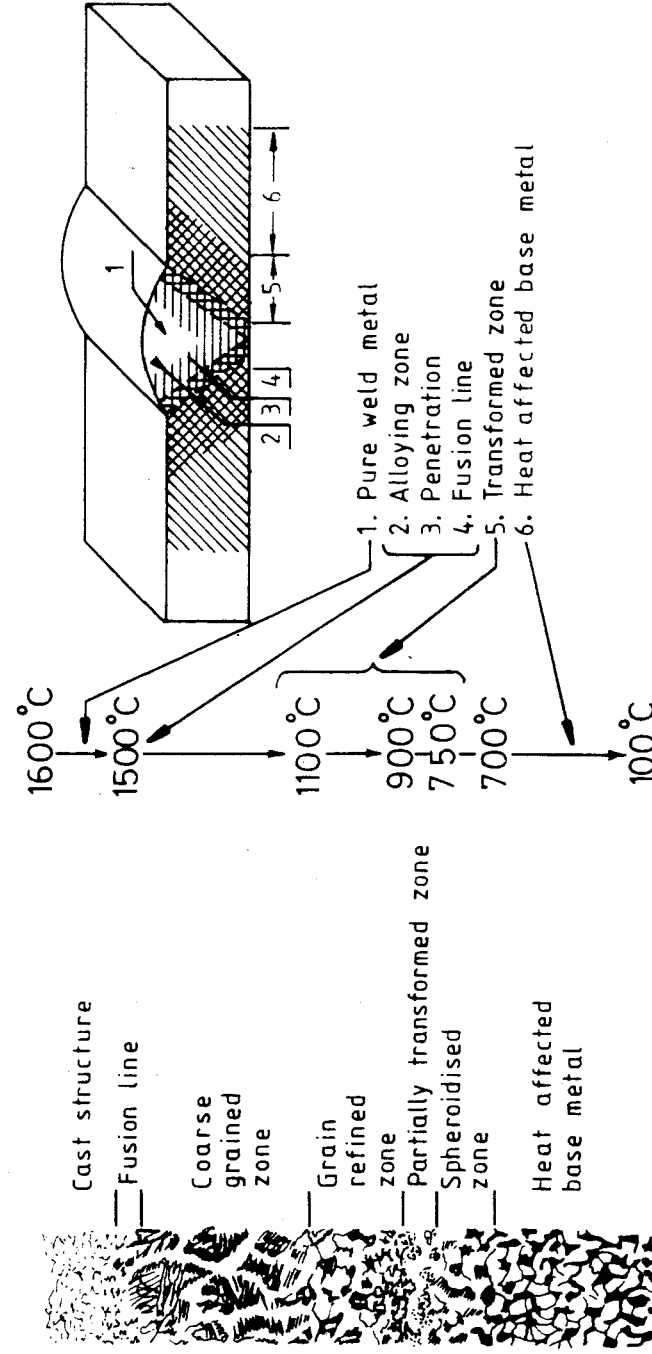


Fig. 5.93 Weld and HAZ structures in fusion welding. (After E. Räsänen and J. Tenkula, *Scandinavian Journal of Metallurgy* 1 (1972) 75.)

precipitates. These precipitates remain fairly stable at temperatures as high as 1500 °C and, at their optimum size of about 10 nm, act as a barrier to grain growth during welding.

Another important problem in welding high-strength steels concerns the formation of martensite. The reason for this is that it is very difficult in welding to avoid the presence of hydrogen. This is because hydrogen-containing compounds are invariably present in fluxes, the electrode material or even in the environment if welding is done outside. In this way atomic hydrogen is absorbed into the molten metal of the fusion weld where it then diffuses rapidly into the heat-affected base metal. If during subsequent cooling, martensite forms, hydrogen (whose solubility in martensite is lower than in ferrite) is forced out of the martensite where it concentrates at the martensite–ferrite phase boundary, or at inclusion boundaries. Thus in combination with weld residual stresses the hydrogen weakens the iron lattice and may initiate cracks. This phenomenon is known as cold cracking. It is found vital in welding to exert a close control over the amount of residual hydrogen in welds and to avoid martensite, particularly in cases where residual stresses may be high. Since it is usually difficult to totally avoid the presence of hydrogen, special CCT diagrams are employed in conjunction with estimated cooling rates in the heat affected material as shown in Fig. 5.94. The essential feature of this type of CCT diagram is that the phase boundaries need to be plotted under conditions of actual welding, or weld simulation, in which both thermal and residual stresses are present²⁰. In the case of weld simulation, a special equipment is employed in which it is possible to programme in the appropriate thermal and stress cycles. As illustrated in Fig. 5.94, the various cooling curves 1–8 represent different heat inputs corresponding to different welding processes or parameters. The parameter t_r in the table refers to the time in seconds for cooling through the temperature range: 800–500 °C, this being almost a constant within the heat-affected zone, and is thus considered a useful parameter in welding in helping to predict microstructure as a function of welding input energy. The working temperature in the table refers to whether or not pre-heating was employed. Thus in Fig. 5.94, martensite is predicted to occur for all welding energies below about 37 500 J/cm (curve 5), this corresponding in practice to a weld deposit on a 20 mm thick plate of the composition given. In practice of course it is more useful if microstructural or cold cracking predictions could be made as a function of differing chemical composition, plate thickness, peak temperature, pre-heating and welding variables. This obviously requires much more complex diagrams than that of Fig. 5.94, and will therefore be correspondingly less accurate, although such diagrams, or welding nomographs as they are called, have been developed for certain applications. The composition variations are estimated using a so-called *carbon equivalent*²¹, in which the effect of the various elements present is empirically expressed as a composition corresponding to a certain carbon content. This is then used to estimate possible martensite formation for the welding conditions given.

Marking	1	2	3	4	5	6	7	8
Heat input J/cm	6500	10000	20000	30000	32500	50000	40000	45000
$t_r(800-500\text{ }^{\circ}\text{C})\text{s}$	2.3	5	10.2	16.5	23.5	32	44.5	57
Working temperature $^{\circ}\text{C}$	Amb	Amb	Amb	Amb	Amb	Amb	200	200

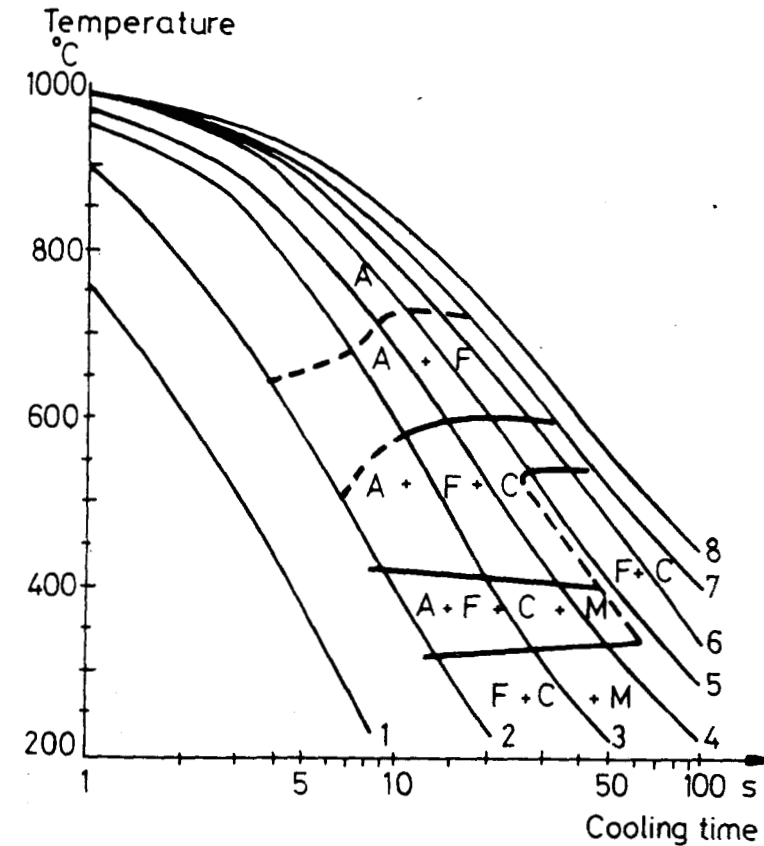


Fig. 5.94 CCT diagrams for a 0.19% C–1.52 Mn–0.55 Si steel, with superimposed cooling rates corresponding to the weld heat inputs given in the table. (From IIW's Doc. 115/11W—382–71, 1971.)

References

1. I.S. Servi and D. Turnbull, *Acta Metallurgica*, **14**:161 (1966).
2. C. Zener, *Journal of Applied Physics*, **20**:950 (1949).
3. G.J Jones and R. Trivedi, *Journal of Applied Physics*, **42**:4299 (1971).
4. C. Laird and H.I. Aaronson, *Acta Metallurgica*, **17**:505 (1969).

5. P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, New York, 1969, p. 92.
6. For a rigorous derivation of this see J. Cahn and W. Hagel in *Decomposition of Austenite by Diffusional Processes*, V. Zackay and H.I. Aaronson (Eds.), p. 134, Interscience, New York, 1962.
7. This treatment is based on a series of papers by J. Cahn. See for example *Acta Metallurgica*, **9**:795 (1961) and *Transactions of the Metallurgical Society of AIME*, **242**:168 (1968).
8. J.W. Martin and R.D. Doherty, *Stability of Microstructure in Metallic Systems*, Cambridge University Press, 1976, Chapter 4.
9. For a treatment of plate-shaped precipitates see M. Ferrante and R.D. Doherty, *Acta Metallurgica*, **27**:1603 (1979).
10. C. Wagner, *Z. Elektrochem.*, **65**:581 (1961), and I.M. Lifshitz and V.V. Slyozov, *J. Phys. Chem. Solids*, **19**:35 (1961).
11. J.M. Rigsbee and H.I. Aaronson, *Acta Metallurgica*, **27**:351–63 and 365–76 (1979).
12. H.I. Aaronson, C. Laird and K.K. Kinsman, *Phase Transformations*, Chapter 8, American Society for Metals, Metals Park, Ohio (1970).
13. D.A. Porter and J.W. Edington, *Proc. R. Soc., Lond.*, **A358**:335–50 (1977).
14. See M. Hillert in 'The Decomposition of Austenite by Diffusional Processes', V.F. Zackay and H.I. Aaronson, (Eds.), Interscience, New York, 1962, p. 197.
15. For a review of the pearlite transformation see M.P. Puls and J.S. Kirkaldy, *Metallurgical Transactions*, **3**:2777 (1972).
16. For detailed description of bainite microstructures see R.F. Hehemann, *Phase Transformations*, American Society for Metals, Metals Park, Ohio, 1970, p. 397.
17. See for example D.E. Coates, *Metallurgical Transactions*, **4**:2313 (1973).
18. R.C. Sharma and G.R. Purdy, *Metallurgical Transactions*, **4**:2303 (1973).
19. B.N.P. Babu, M.S. Bhat, E.R. Parker and V.F. Zackay, *Metallurgical Transactions*, **7A**:17 (1976).
20. See, e.g. H.C. Cotton, 'Material requirements for offshore structures', Rosenhain Centenary Conference, p. 19, The Metals Society and Royal Society, (1975).
21. From IIW's Doc. IIS/IIW-382-71, 'Guide to the weldability of C–Mn steels and C–Mn microalloyed steels', (1971).
22. H.B. Aaron and H.I. Aaronson, *Acta Metallurgica*, **16**:789 (1968).

Further Reading

H.I. Aaronson (Ed.), *Lectures on the Theory of Phase Transformations*, Metallurgical Society of AIME, 1975.

- H.K.D.H. Bhadeshia and J.W. Christian, 'Bainite in steels', *Metallurgical Transactions, A*, **21A**:767 (1990).
- J. Burke, *The Kinetics of Phase Transformations in Metals*, Pergamon Press, Oxford, 1965.
- G.A. Chadwick, *Metallography of Phase Transformations*, Chapters 6 and 7, Butterworths, London, 1972.
- J.W. Christian, *The Theory of Transformations in Metals and Alloys*, 2nd edition, Pergamon, 1975, Chapters 10, 11, 12.
- R.D. Doherty, 'Diffusive phase transformations in the solid state', Chapter 14 in *Physical Metallurgy*, R.W. Cahn and P. Haasen (Eds.), North-Holland, 1983.
- K.E. Easterling, *Introduction to the Physical Metallurgy of Welding*, Butterworth-Heinemann, London, 2nd edn, 1992.
- M.E. Fine, *Introduction to Phase Transformations in Condensed Systems*, Macmillan, New York, 1964.
- R.W.K. Honeycombe, *Steels—Microstructure and Properties*, Edward Arnold, London, 1980.
- E. Hornbogen, 'Physical metallurgy of steels', Chapter 16 in *Physical Metallurgy*, R.W. Cahn and P. Haasen (Eds.), North-Holland, 1983.
- A. Kelly and R.B. Nicholson, 'Precipitation hardening', *Progress in Materials Science*, Vol. **10**, p. 149–392, Macmillan, New York, 1963.
- A. Kelly and R.B. Nicholson (Eds.), *Strengthening Methods in Crystals*, Applied Science, London, 1971.
- J.W. Martin, *Precipitation Hardening*, Pergamon, Oxford, 1968.
- I. Polmear, *Light Alloys*, 2nd edn, Edward Arnold, 1989.
- K.C. Russell and H.I. Aaronson (Eds.), *Precipitation Processes Solids*, The Metallurgical Society of AIME, 1978.
- I. Tamura, C. Ouchi, T. Tanaka and H. Sekine, *Thermomechanical Processing of High Strength Low Alloy Steels*, Butterworths, London, 1988.
- V.F. Zackay and H.I. Aaronson, (Eds.) *Decomposition of Austenite by Diffusional Process*, Interscience, New York, 1962.
- Phase Transformations* American Society for Metals, Metals Park, Ohio, 1970.
- Phase Transformations* Institution of Metallurgists, London, 1979.

Exercises

- 5.1 An approximate expression for the total driving force for precipitation in a regular solution (ΔG_0 in Fig. 5.3) is

$$\Delta G_0 = RT \left[X_0 \ln \frac{X_0}{X_e} + (1 - X_0) \ln \frac{(1 - X_0)}{(1 - X_e)} \right] - \Omega (X_0 - X_e)^2$$

where X_0 and X_e are the mole fractions of solute defined in Fig. 5.3.

(a) Use this equation to estimate the total free energy released when

- $\alpha' \rightarrow \alpha + \beta$ at 600 K if $X_0 = 0.1$, $X_e = 0.02$ and $\Omega = 0$ (ideal solution) ($R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$).
- (b) Estimate the volume fraction of precipitate at equilibrium if β is pure solute ($X_B^\beta = 1$). (Assume the molar volume is constant.)
- (c) If the alloy is heat treated to produce a precipitate dispersion with a spacing of 50 nm estimate the total α/β interfacial area m^{-2} of alloy. (Assume a simple cubic array.)
- (d) If $\gamma_{\alpha\beta} = 200 \text{ mJ m}^{-2}$ what is the total interfacial energy m^{-2} of alloy? mol^{-1} of alloy? ($V_m = 10^{-5} \text{ m}^3$).
- (e) What fraction of the total driving force would remain as interfacial energy in the above case?
- (f) Repeat c–e for a dispersion of 1 μm spacing.
- 5.2 Use the methods of Chapter 1 to derive the expression for ΔG_0 in problem 5.1.
- 5.3 In dilute or ideal solutions the driving force for precipitate nucleation (assuming $X_B^\beta = 1$) is given approximately by

$$\Delta G_n = RT \ln \frac{X_0}{X_e} \text{ per mole of precipitate}$$

where X_0 and X_e are the mole fractions of solute defined in Fig. 5.3.

- (a) Evaluate ΔG_n for the precipitate in problem 5.1.
- (b) Assuming homogeneous nucleation, what will be the critical nucleus radius?
- (c) How does the mean precipitate size in problem 5.1c compare with the size of the critical nucleus?
- 5.4 Derive the expression for ΔG_n in problem 5.3. (Use equation 1.68.)
- 5.5 (a) Calculate θ in Fig. 5.6 if $\gamma_{\alpha\beta} = 500$ and $\gamma_{\alpha\alpha} = 600 \text{ mJ m}^{-2}$.
 (b) Evaluate the magnitude of the shape factor $S(\theta)$ for this nucleus.
- 5.6 Imagine the Fe–0.15 wt% C alloy in Fig. 5.45 is austenitized above A_3 , and then quenched to 800 °C where ferrite nucleates and covers the austenite grain boundaries.
- (a) Draw a composition profile normal to the α/γ interface after partial transformation assuming diffusion-controlled growth.
- (b) Derive an approximate expression for the thickness of the ferrite slabs as a function of time.
- (c) Given that $D_\gamma^\gamma(800^\circ\text{C}) = 3 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$ plot the thickness as a function of time.
- (d) If the austenite grain size is 300 μm extend the above curve to long times. (State any simplifying assumptions you make.)
- 5.7 Derive Equation 5.34.
- 5.8 (a) By considering short transformation times derive expressions for k and n in Equation 5.39 for the pearlite transformation when nucleation is restricted to grain corners and all nuclei form at time zero (site saturation). Assume spherical pearlite nodules and a cubic grain structure with a cube side d .

- (b) Repeat the above for grain-boundary nucleation again assuming site saturation. In this case pearlite grows as grain-boundary slabs.
- 5.9 Draw schematic diagrams to show how growth rate and nucleation rate should vary with temperature for civilian transformations that are induced by an increase in temperature.
- 5.10 A and B form a regular solution with a positive heat of mixing so that the A–B phase diagram contains a miscibility gap.
- (a) Starting from Equation 1.39 derive an equation for d^2G/dX_B^2 , assuming $G_A = G_B = 0$.
- (b) Use the above equation to calculate the temperature at the top of the miscibility gap T_c in terms of Ω .
- (c) Plot the miscibility gap for this system. Hint: the limits of solubility for this simple case are given by $dG/dX_B = 0$.
- (d) On the same diagram plot the locus of $d^2G/dX_B^2 = 0$, i.e. the chemical spinodal.
- 5.11 By expressing G as a Taylor series, i.e.

$$G(X_0 + \Delta X) = G(X_0) + \frac{dG}{dX}(\Delta X) + \frac{d^2G}{dX^2} \frac{(\Delta X)^2}{2} + \dots$$

show that Equation 5.43 is valid for small values of ΔX .

- 5.12 How should alloy composition affect the initial wavelength of a spinodally decomposed microstructure at a given temperature?
- 5.13 (a) Account for the location of massive transformations in Table 3.5.
 (b) Why do massive transformations generally occur at lower temperatures but higher rates than precipitation transformations?

6

Diffusionless Transformations

One of the most important technological processes is the hardening of steel by quenching. If the steel is quenched rapidly enough from the austenitic field, there is insufficient time for eutectoidal diffusion-controlled decomposition processes to occur, and the steel transforms to martensite—or in some cases martensite with a few per cent of retained austenite. This transformation is important and best known in connection with certain types of stainless steels, quenched and tempered steels and ball bearing steels. Important recent developments involving the martensitic transformation in steels include maraging steels (precipitation-hardened martensite), TRIP steels (transformation induced by plastic deformation), ausforming steels (plastically deformed austenite prior to quenching) and dual phase steels (a mixture of ferrite + martensite obtained by quenching from the $\gamma + \alpha$ field).

Because of the technological importance of hardened steel we shall mainly be concerned with this transformation, although martensite is a term used in physical metallurgy to describe any diffusionless transformation product, i.e. any transformation in which from start to completion of the transformation individual atomic movements are less than one interatomic spacing. The regimented manner in which atoms change position in this transformation has led to it being termed *military*, in contrast to diffusion-controlled transformations which are termed *civilian*. In principle, all metals and alloys can be made to undergo diffusionless transformations provided the cooling rate or heating rate is rapid enough to prevent transformation by an alternative mechanism involving the diffusional movement of atoms. Martensitic transformations can thus occur in many types of metallic and non-metallic crystals, minerals and compounds. In the case of martensite in steel, the cooling rate is such that the majority of carbon atoms in solution in the fcc γ -Fe remain in solution in the α -Fe phase. Steel martensite is thus simply a supersaturated solid solution of carbon in α -Fe. The way in which this transformation occurs, however, is a complex process and even today the transformation mechanism, at least in steels, is not properly understood. The main purpose of this chapter is to consider some of the characteristics of martensitic transformations including a brief study of their crystallography, and to examine possible theories of how the phase nucleates and grows. We shall then consider the process of tempering steel martensites and finally give some examples of engineering materials based on martensitic transformations.

6.1 Characteristics of Diffusionless Transformations

There have been a number of excellent reviews of martensitic transformations, and the most complete treatments to date have been given by Christian (1965)¹ and Nishiyama (1978)². The formation of martensite appears from micrographs to be a random process and the way it is observed to develop is illustrated schematically in Fig. 6.1a and b. As seen from Fig. 6.1a, the martensitic phase (designated α') is often in the shape of a lens and spans initially an entire grain diameter. The density of plates does not appear to be a function of the grain size of the austenite. For example it is observed to form randomly throughout a sample with a plate density which appears to be independent of grain size. Where the plates intersect the surface of a polished specimen they bring about an elastic deformation, or tilting of the surface as shown in Fig. 6.2. Observations have shown that, at least macroscopically, the transformed regions appear coherent with the surrounding austenite. This means that intersection of the lenses with the surface of the specimen does not result in any discontinuity. Thus, lines on a polished surface are displaced, as illustrated in Fig. 6.2a, but remain *continuous* after the transformation. It has been shown that a fully grown plate spanning a whole grain may form within $\sim 10^{-7}$ s which means that the α'/γ interface reaches almost the speed of sound in the solid. Martensite is thus able to grow independently of thermal activation, although some Fe-Ni alloys do exhibit isothermal growth characteristics. This great speed of formation makes martensite nucleation and growth a difficult process to study experimentally.

It is seen in Fig. 6.1a and b that the volume fraction of martensite increases by the systematic transformation of the austenite remaining between the plates that have already formed. The first plates form at the M_s (*martensite start*) temperature. This temperature is associated with a certain driving force for the diffusionless transformation of γ into α' as shown in Fig. 6.3a and b. In low-carbon steels, $M_s \approx 500^\circ\text{C}$ (Fig. 6.3c), but increasing C contents progressively decrease the M_s temperature as shown. The M_f temperature (*martensite finish*) corresponds to that temperature below which further cooling does not increase the amount of martensite. In practice the M_f may not correspond to 100% martensite, and some retained austenite can be left even below M_f . The retention of austenite in such cases may be due to the high elastic stresses between the last martensite plates to form, which tend to suppress further growth or thickening of existing plates. As much as 10–15% retained austenite is a common feature of especially the higher C content alloys such as those used for ball bearing steels. Figure 6.3d is a TTT diagram used for estimating the speed of quench necessary to obtain a given microstructure. These diagrams are plotted and used in technological applications for any *one* particular alloy, and that illustrated for example applies to only one carbon content, as shown.

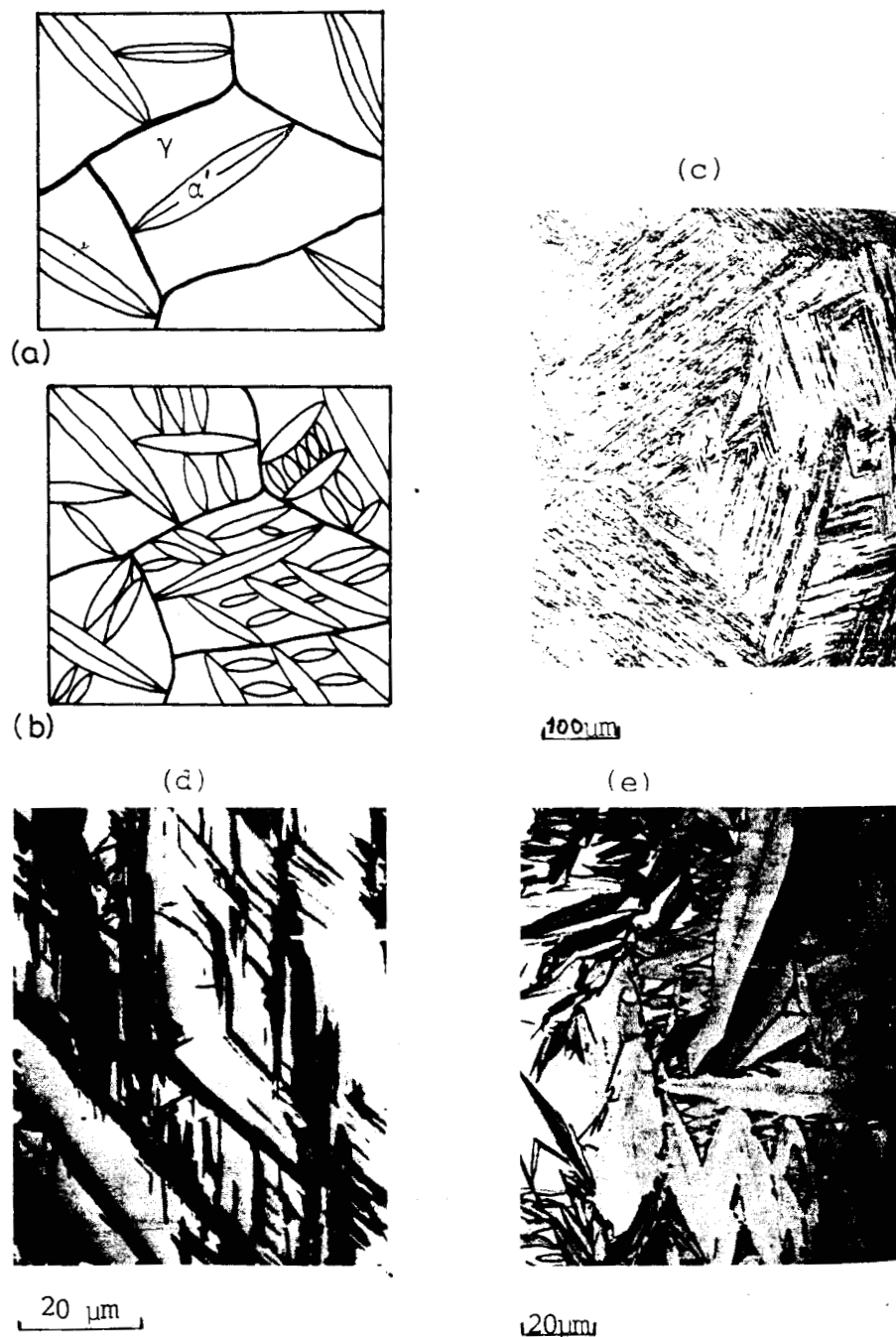


Fig. 6.1 (a), (b) Growth of martensite with increasing cooling below M_s . (c)–(e) Different martensite morphologies in iron alloys: (c) low C (lath). (d) medium C (plate). (e) Fe–Ni (plate).

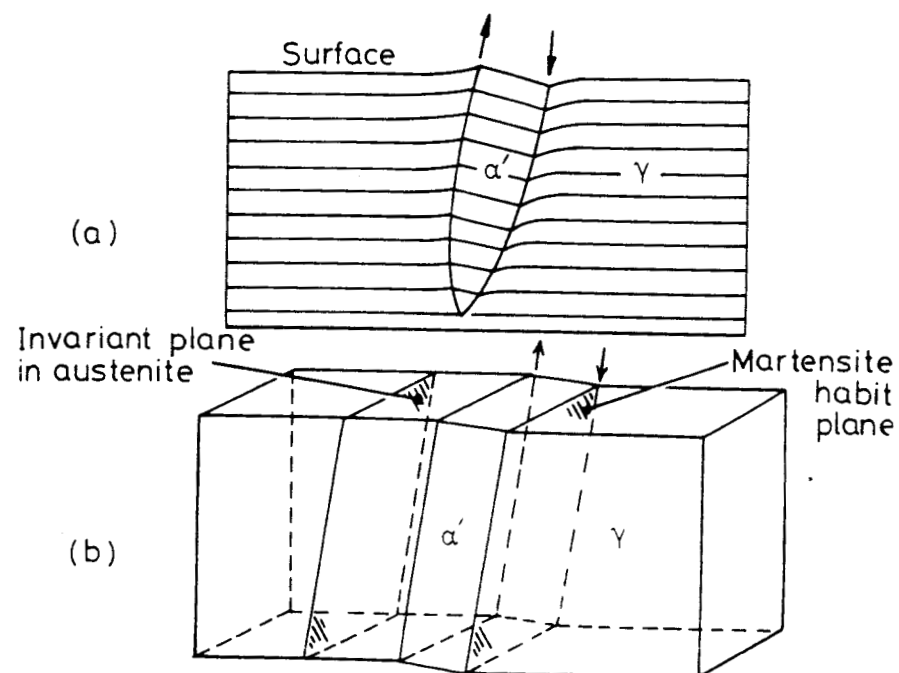


Fig. 6.2 Illustrating how a martensite plate remains (macroscopically coherent with the surrounding austenite and even the surface it intersects.

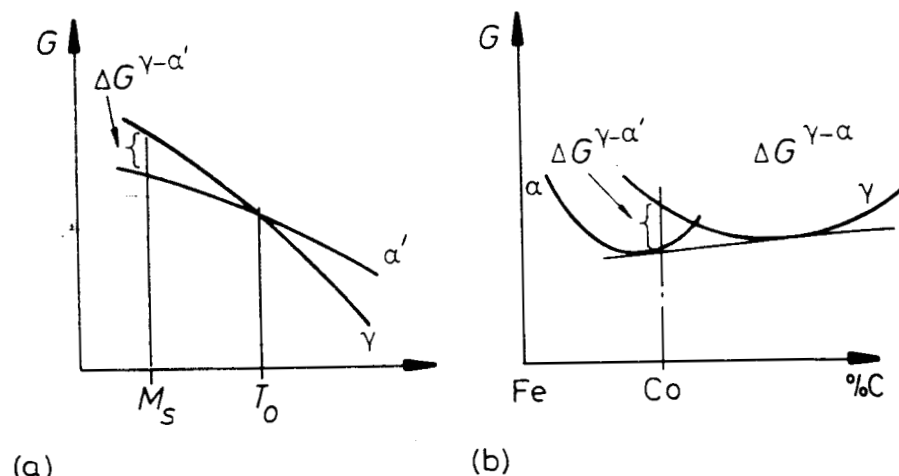
By analogy with Equation 1.17, the driving force for the nucleation of martensite at the M_s temperature should be given by:

$$\Delta G^{\gamma \rightarrow \alpha'} = \Delta H^{\gamma \rightarrow \alpha'} \frac{(T_0 - M_s)}{T_0} \quad (6.1)$$

where T_0 and M_s are defined in Fig. 6.3a. Some calorimetric measurements of ΔH are given in Table 6.1 for a number of alloys exhibiting martensitic transformations, together with the corresponding amounts of undercooling and free energy changes. Note especially in this table the large differences in $\Delta G^{\gamma \rightarrow \alpha'}$ between ordered and disordered alloys, the ordered alloys exhibiting a relatively small undercooling. We shall now examine the atomic structures of steel austenite and martensite in more detail.

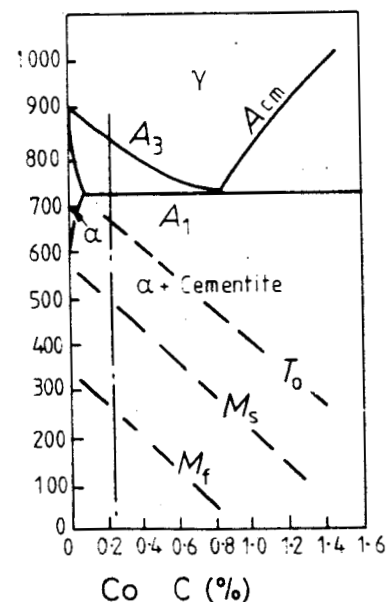
6.1.1 The Solid Solution of Carbon in Iron

In an fcc (or hcp) lattice structure, there are two possible positions for accommodating interstitial atoms as shown in Fig. 6.4. These are: the tetrahedral site which is surrounded by four atoms and the octahedral site which has six nearest neighbours. The sizes of the largest atoms that can be accommodated in these holes without distorting the surrounding matrix atoms can be calculated if it is assumed that the atoms are close-packed hard



(a)

(b)



(c)

(d)

Fig. 6.3 Various ways of showing the martensite transformation. (a) Free energy-temperature diagram for austenite and martensite of fixed carbon concentration (C_0 in (b)). (b) Free energy-composition diagram for the austenite and martensite phases at the M_s temperature. (c) Iron-carbon phase diagram with T_0 as defined in (a). M_s and M_f superimposed. (d) M_s and M_f in relation to the TTT diagram for alloy C_0 in (c).

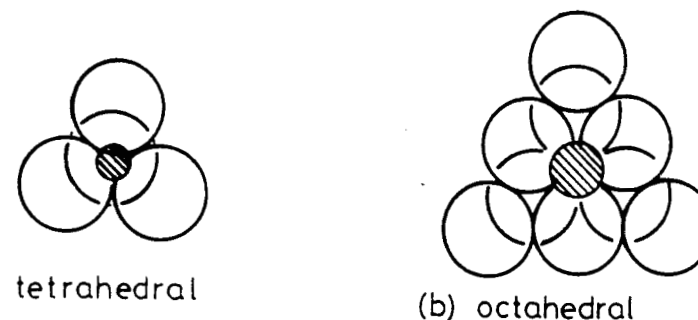


Fig. 6.4 Illustrating possible sites for interstitial atoms in the fcc or hcp lattices.

spheres. Such a calculation gives:

$$\text{tetrahedral interstice } d_4 = 0.225 D$$

$$\text{octahedral interstice } d_6 = 0.414 D \quad (6.2a)$$

where D is the diameter of the parent atoms and d_4 and d_6 are the maximum interstitial diameters in the two types of site. In the case of γ -iron, at ambient temperature $D = 2.52 \text{ \AA}$, so that interstitial atoms of diameter $0.568 \text{ \AA}$ or $1.044 \text{ \AA}$ can be contained in tetrahedral and octahedral interstices without distorting the lattice. However, the diameter of a carbon atom is $1.54 \text{ \AA}$. This means that considerable distortion of the austenite lattice must occur to contain carbon atoms in solution and that the octahedral interstices should be the most favourable.

The possible positions of interstitials in the bcc lattice are shown in Fig. 6.5a. It is seen that there are three possible octahedral positions ($\frac{1}{2}[100]$),

Table 6.1 Comparisons of Calorimetric Measurements of Enthalpy and Undercooling in some martensitic alloys

G. Guénin, Ph.D. thesis, Polytechnical Inst. of Lyon, 1979

Alloy	$\Delta H^{\gamma \rightarrow \alpha'}$ (J mol ⁻¹)	$T_0 - M_s$ (K)	$-\Delta G^{\gamma \rightarrow \alpha'}$ (J mol ⁻¹)
Ti-Ni	1550	20	92
Cu-Al	170-270	20-60	19.3 ± 7.6
Au-Cd	290	10	11.8
Fe-Ni 28%	1930	140	840
Fe-C			1260
Fe-Pt 24%	340	10	17
ordered			
Fe-Pt	2390	~ 150	~ 1260
disordered			

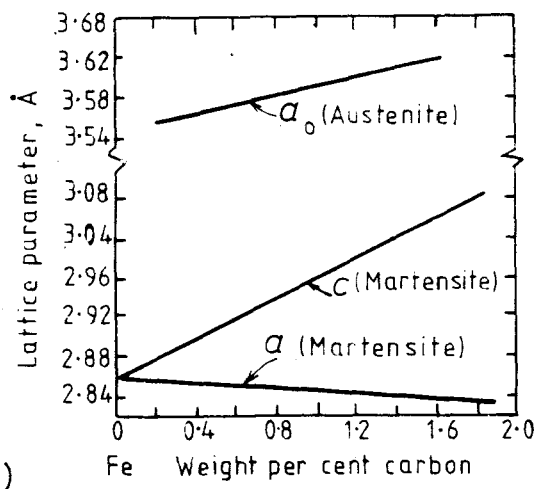
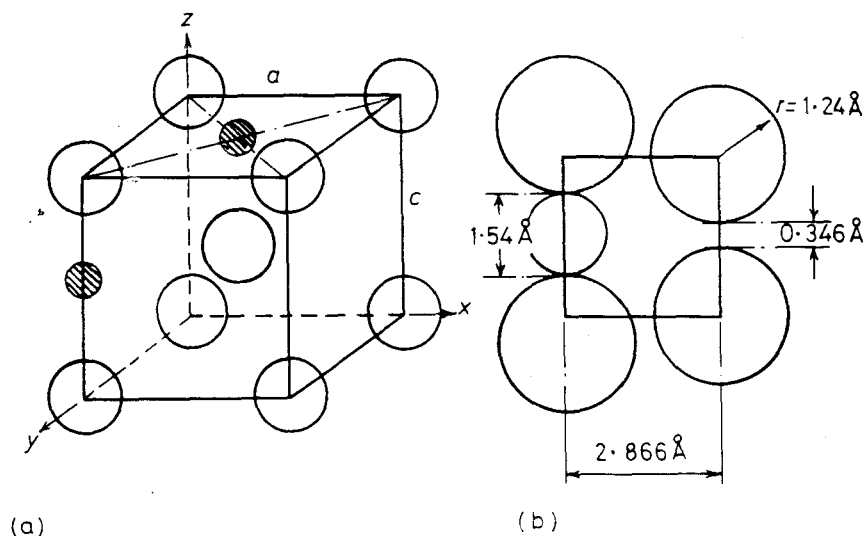


Fig. 6.5 Illustrating (a) possible sites for interstitial atoms in bcc lattice, and (b) the large distortion necessary to accommodate a carbon atom (1.54 Å diameter) compared with the space available (0.346 Å). (c) Variation of a and c as a function of carbon content. (After C.S. Roberts, *Transactions AIME* 191 (1953) 203.)

$\frac{1}{2}[010]$, $\frac{1}{2}[001]$), and six possible tetrahedral spaces for each unit cell. In this case, the maximum sizes of interstitials that can be accommodated without distorting the lattice are as follows:

$$\begin{aligned} d_4 &= 0.291 D \\ d_6 &= 0.155 D \end{aligned} \quad (6.2b)$$

The interesting feature of the bcc lattice is that although there is more 'free' space than the close-packed lattices, the larger number of possible interstitial positions means that the space available *per interstitial* is less than for the fcc structure (compare Equations 6.2a and 6.2b). In spite of the fact that $d_6 < d_4$, measurements of carbon and nitrogen in solution in iron show that these interstitials in fact prefer to occupy the octahedral positions in the bcc lattice. This causes considerable distortion to the bcc lattice as illustrated in Fig. 6.5b. It is conjectured that the bcc lattice is weaker in the $\langle 100 \rangle$ directions due to the lower number of near and next nearest neighbours compared to the tetrahedral interstitial position (see, e.g. Cottrell, 1963)³. The estimated atomic diameters of pure carbon and nitrogen are 1.54 and 1.44 Å respectively, although these values are very approximate. It should also be remembered that in a given steel relatively few $\langle \frac{1}{2}00 \rangle$ sites are occupied. Nevertheless, the martensitic Fe-C lattice is distorted to a bct structure as shown in Fig. 6.5c. These measurements, made by X-ray diffraction at -100°C to avoid carbon diffusion, show that the c/a ratio of the bct lattice is given by:

$$c/a = 1.005 + 0.045(\text{wt\% C}) \quad (6.3)$$

As seen by these results, the distortion of the lattice in one direction (z) causes a contraction in the two directions normal to z (x, y). In fact, these measurements suggest a certain long-range order in the distribution of the carbon interstitials.

6.2 Martensite Crystallography⁴

A feature of the microstructures shown in Fig. 6.1 is the obvious crystallographic dependence of martensite plate formation. Within a given grain, all the plates grow in a limited number of orientations. In the case of iron alloys, for example, the orientation variants and even plate morphology chosen turn out to be dependent upon alloy content, particularly carbon or nickel, as illustrated in Table 6.2.

The irrational nature of the growth planes of high carbon or high nickel martensites has been the subject of much discussion in the literature for the following reason: if martensite is able to grow at speeds approaching the speed of sound, then some sort of highly mobile dislocation interface is required. The problem is then to explain the high mobility of an interface moving on austenite planes *not* always associated with dislocation glide. Yet another is that the growth or *habit* plane of martensite is observed to be *macroscopically undistorted*, i.e. the habit plane is a plane which is common to both the austenite and martensite in which all directions and angular separations in the plane are unchanged during the transformation. That this is so can be reasoned in conjunction with Fig. 6.2. The absence of plastic deformation in the form of a discontinuity at the surface shows that the shape strain does not cause any significant rotation of the habit plane. If the habit plane had been rotated, plastic deformation would be necessary to maintain

Table 6.2 Orientation Relationships and Habit Planes of some Martensites

Data mainly from Z. Nishiyama, *Martensitic Transformation*, Academic Press, pp. 115-123 (1978).

Alloy	Transformation	Orientation relation	Habit plane	Defect structure
Fe-(0-0.4 wt% C)	fcc $\rightarrow$ bct	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{111\}_\gamma$	needles or laths, high dislocation density
Fe-(0.5-1.4 wt% C)	fcc $\rightarrow$ bct	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{225\}_\gamma$	mixed twins and dislocations
Fe-(1.5-1.8 wt% C)	fcc $\rightarrow$ bct	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{259\}_\gamma$	mainly twinned
Fe-(27-34 wt% Ni)	fcc $\rightarrow$ bcc	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{259\}_\gamma$	mainly twinned
Fe-(11-29% Ni)-(0.4-1.2% C)	fcc $\rightarrow$ bct	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{259\}_\gamma$	mainly twinned
Fe-(7-10% Al)-2% C	fcc $\rightarrow$ bct	$(111)_\gamma // (011)_\alpha$ $[10\bar{1}]_\gamma // [111]_\alpha$	$\{3\ 10\ 15\}_\gamma$	twinned
Fe-(2.8-8% Cr)-(1.1-1.5% C)	fcc $\rightarrow$ bct	probably as for Fe-C alloys	$\{225\}_\gamma$	mixed twins and dislocations
Fe-4.5 wt% Cu	fcc $\rightarrow$ bcc	as for Fe-C alloys	$\{112\}_\gamma$	laths, dislocations stacking faults
Co	fcc $\rightarrow$ hcp	$(111)_\gamma // (0001)_\alpha$ $[112]_\gamma // [1100]_\alpha$ or $[110]_\gamma // [1120]_\alpha$	$\{111\}_\gamma$	
Ti	bcc $\rightarrow$ hcp	$(110)_\gamma // (0001)_\alpha$	$\{89, 12\}_\gamma$ $\{133\}_\gamma$	mixed twins and dislocations
Zr	bcc $\rightarrow$ hcp bcc $\rightarrow$ hcp	$(110)_\gamma // (0001)_\alpha$ $(110)_\gamma // (0001)_\alpha$	$\{569\}_\gamma$ $\{145\}_\gamma$ $\{441\}$	—

coherence between the martensite and parent austenite and this would have resulted in additional displacements of the surface, or of the lines traversing the plate. In order that the habit plane is left undistorted, the martensitic transformation appears to occur by a homogeneous shear parallel to the habit plane (see Fig. 6.2). Since the $\gamma \rightarrow \alpha'$ transformation is also associated with $\sim 4\%$ expansion, this implies in turn that the dilatation in question must take place normal to the habit plane, i.e. normal to the lens. However, some homogeneous dilatation of the habit plane may be necessary.

The question now arises: can the bct martensite lattice structure be generated by simple shear parallel to the habit plane, together with a small dilatation normal to the plane? In order to answer this question adequately we must consider the crystallography of the $\gamma \rightarrow \alpha'$ transformation in more detail.

It has been stated that the habit plane of a martensite plate remains undistorted following the transformation. An analogous situation is found in twinning as illustrated in Fig. 6.6a and b. It is convenient to consider the $(111)_\gamma$ $(112)_\gamma$ twinning reaction illustrated in Fig. 6.6a in terms of the homogeneous shear of a sphere, Fig. 6.6b. In the shearing plane K , the lattice is undistorted, i.e. it is *invariant*. Let us assume first that the equivalent macroscopic shape change in the formation of a martensite plate is a twinning shear occurring parallel to the habit (or twinning) plane, plus a simple uniaxial tensile dilatation perpendicular to the habit plane. A strain of this type is termed: an *invariant plane strain*, because a shear parallel to the habit plane, or an extension or contraction perpendicular to it, cannot change the positions or magnitude of vectors lying in the plane. We shall now try to answer the question of whether the fcc lattice can be homogeneously deformed to generate the bct structure.

6.2.1 The Bain Model of the fcc $\rightarrow$ bct Transformation

In 1924, Bain⁵ demonstrated how the bct lattice could be obtained from the fcc structure with the minimum of atomic movement, and the minimum of strain in the parent lattice. To illustrate this we shall use the convention that x, y, z and x', y', z' represent the original and final axes of the fcc and bcc unit cells as illustrated in Fig. 6.7. As shown by this figure, an elongated unit cell of the bcc structure can be drawn within two fcc cells. Transformation to a bcc unit cell is achieved by: (a) contracting the cell 20% in the z direction and expanding the cell by 12% along the x and y axes. In the case of steels, the carbon atoms fit into z' axes of the bcc cell at $\frac{1}{2}(100)$ positions causing the lattice to elongate in this direction. In 1 atomic % C steel, for example, carbon occupies one position along the z' axis for every 50 iron unit cells. The positions occupied by the carbon atoms in the bct structure do not exactly correspond to the equivalent octahedral positions in the parent fcc structure, and it is assumed that small shuffles of the C atoms must take place during the transformation.

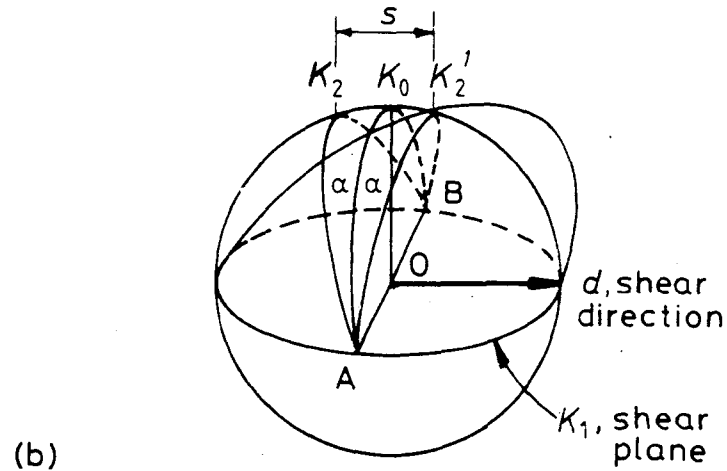
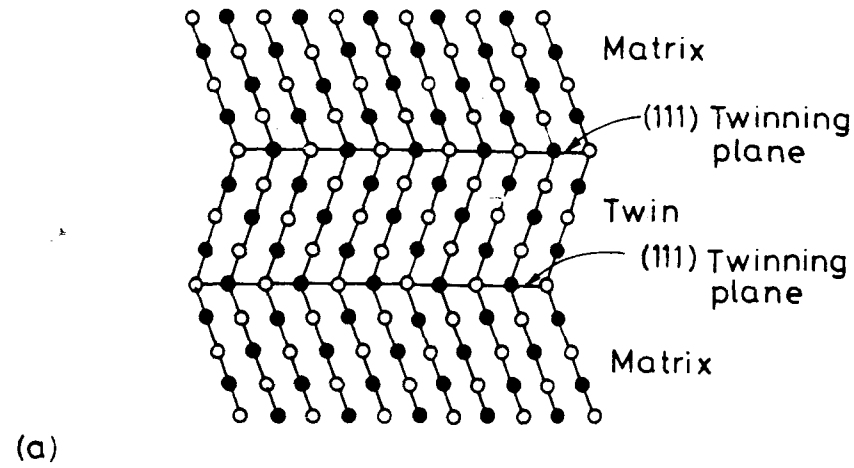


Fig. 6.6 (a) Showing the twinning of an fcc structure. Black and white circles represent atoms on different levels. (R.E. Reed-Jr., *Physical Metallurgy Principles*, 2nd edn., Van Nostrand, 1973.) (b) Graphical representation of a twinning shear occurring on a plane K_1 in a direction d (from C.M. Wayman, *Introduction to the Crystallography of Martensite Transformations*, MacMillan, New York, 1964).

It is an interesting fact that the Bain deformation involves the absolute minimum of atomic movements in generating the bcc from the fcc lattice. Examination of Fig. 6.7 shows that the Bain deformation results in the following correspondence of crystal planes and directions:

$$\begin{aligned} (111)_\gamma &\rightarrow (011)_{\alpha'} \\ [\bar{1}01]_\gamma &\rightarrow [\bar{1}\bar{1}1]_{\alpha'} \\ [1\bar{1}0]_\gamma &\rightarrow [100]_\alpha \\ [11\bar{2}]_\gamma &\rightarrow [01\bar{1}]_\alpha \end{aligned}$$

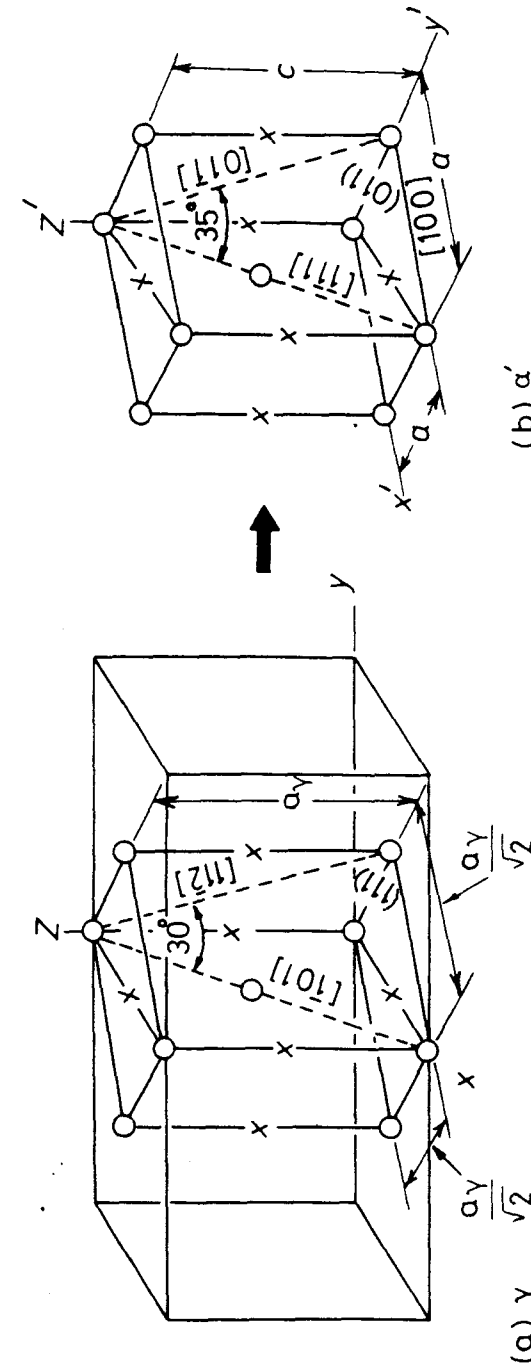


Fig. 6.7 Bain correspondence for the $\alpha \rightarrow \alpha'$ transformation. Possible interstitial sites for carbon are shown by crosses. To obtain α' the γ unit cell is contracted about 20% on the C axis and expanded about 12% on the a axes.

Experimental observations of orientation relationships between austenite and martensite show that $[111]_{\gamma}$ planes are approximately parallel to $[011]_{\alpha}$ planes, and that the relative directions can vary between $\langle 101 \rangle_{\gamma} \parallel \langle 1\bar{1}1 \rangle_{\alpha}$ (the Kurdjumov – Sachs relation) and $\langle 1\bar{1}0 \rangle_{\gamma} \parallel \langle 101 \rangle_{\alpha}$ (the Nishiyama-Wasserman relation). These two orientations differ by $\sim 5^\circ$ about $[111]_{\gamma}$.

By using the sphere $\rightarrow$ ellipsoid transformation applied earlier to demonstrate the twinning shear (Fig. 6.6) we can now test whether the Bain deformation also represents a pure deformation in which there is an undeformed (invariant) plane. If a sphere of unit radius represents the fcc structure then after the Bain distortion it will be an ellipsoid of revolution with two axes (x' and y') expanded by 12% and the third axis (z') contracted by 20%. The $x'z'$ section through the sphere before and after distortion is shown in Fig. 6.8. In this plane the only vectors that are not shortened or elongated by the Bain distortion are OA or $O'A'$. However in order to find a *plane* in the fcc structure that is not distorted by the transformation requires that the vector OY' (perpendicular to the diagram) must also be undistorted. This is clearly not true and therefore the Bain transformation does not fulfil the requirements of bringing about a transformation with an undistorted plane.

Hence the key to the crystallographic theory of martensitic transformations is to postulate an additional distortion which, in terms of Fig. 6.8, reduces the extension of y' to zero (in fact a slight rotation, θ , of the AO plane should also be made as shown in the figure). This second deformation can be in the form of dislocation slip or twinning as illustrated in Fig. 6.9. Applying the twinning analogy to the Bain model, we can see that an internally twinned martensite plate can form by having alternate regions in the austenite undergo the Bain strain along different contraction axes such that the net distortions are compensated. By also adjusting the width of the individual twins, the habit plane of the plate can even be made to adopt any desired orientation. These features of twinned martensite plates are illustrated in Fig. 6.10. In this figure ϕ defines the angle between some reference plane in the austenite and the martensite habit plane. It is seen that ϕ is a function of twin widths I, II (see,

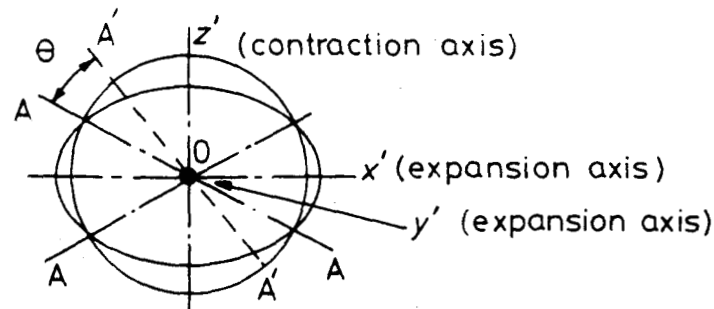


Fig. 6.8 The Bain deformation is here simulated by the pure deformation in compressing a sphere elastically to the shape of an oblate ellipsoid. As in the Bain deformation, this 'transformation' involves two expansion axes and one contraction axis.

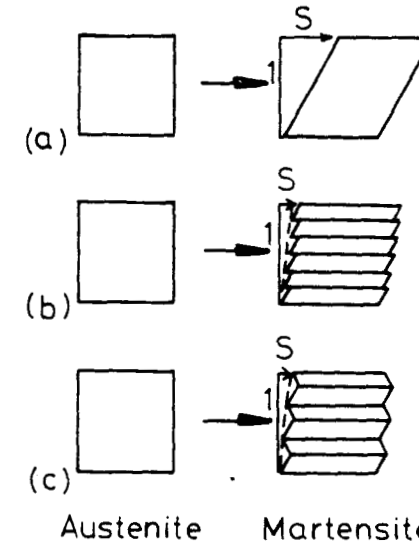


Fig. 6.9 This figure illustrates schematically how dislocation glide or twinning of the martensite can compensate for a pure lattice deformation such as a Bain deformation and thereby reduce the strain of the surrounding austenite. The transformation shear (s) is defined. Note how s can be reduced by slip or twinning.

e.g. Fig. 6.9c). On this basis, the habit plane of the martensite plate can be defined as a plane in the austenite which undergoes no *net (macroscopic) distortion*. By 'net distortion', it is meant that the distortion when averaged over many twins is zero. There will of course be local regions of strain energy associated with the α'/γ interface of the twins at the edge of the plate. However, if the plate is very thin (a few atomic spacings) this strain can be relatively small.

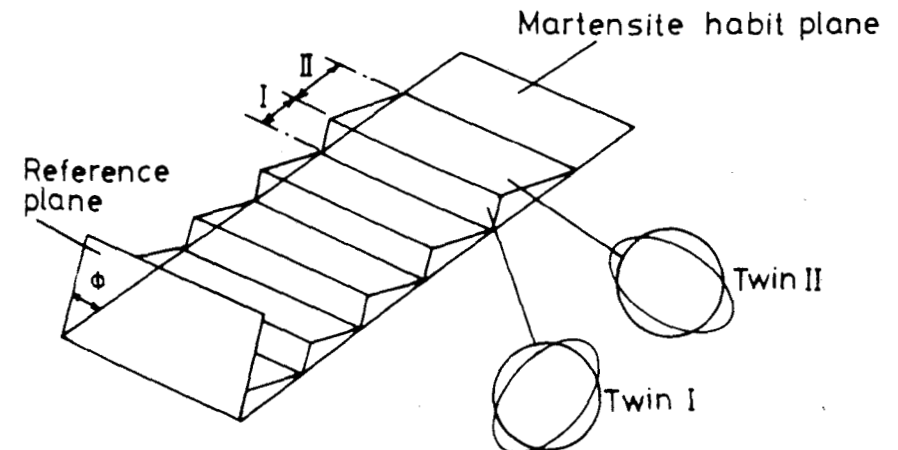


Fig. 6.10 Twins in martensite may be self-accommodating and reduce energy by having alternate regions of the austenite undergo the Bain strain along different axes.

In the crystallographic theory, it is assumed that slip or twinning occurs on suitable $\{111\}\{112\}_\alpha$ systems, corresponding to equivalent $\{1\bar{1}0\}\{110\}_\gamma$ planes in the austenite. Since the $\{112\}_\alpha\{111\}_\alpha$ system is that commonly adopted for bcc slip or twinning, the physical requirements of the theory are satisfied.

6.2.2 Comparison of Crystallographic Theory with Experimental Results

Some plots of experimental measurements typical of habit planes in steel martensites are shown in Fig. 6.11. These results indicate that there is a fairly wide scatter in experimental measurements for a given type of steel, and that alloying additions can have a marked effect on the habit plane. It appears that on reaching a critical carbon content, martensite in steel changes its habit plane, these transitions being approximately $\{111\} \rightarrow \{225\} \rightarrow \{259\}_\gamma$ with increasing C content (there is overlap of these transitions in practice). As a general rule, the $\{111\}$ martensites are associated with a high dislocation density lath morphology, or consist of bundles of needles lying on $\{111\}_\gamma$ planes, while the $\{225\}_\gamma$ and $\{259\}_\gamma$ martensites have a mainly twinned plate or lens morphology. However, any exact morphological description of martensite is not possible since, after thickening and growth, the shapes of the martensites are often quite irregular. Twinning is more predominant at high carbon or nickel contents and is virtually complete for $\{259\}_\gamma$ martensites. In stainless steel, the habit plane is thought to be nearer $\{11\bar{2}\}$, which has been explained in terms of a lattice invariant shear on $\{101\}\{10\bar{1}\}_\alpha$, corresponding to $\{111\}\{1\bar{1}2\}_\gamma$. Transmission electron micrographs of lath and twinned steel martensites are shown in Fig. 6.12, which also illustrates the 'classical' definition of lath and plate morphologies.

The notable success of the crystallographic theory is that it was able to predict the fine substructure (twinning or slip) of martensite before it was actually observed in the electron microscope. For a typical (high-carbon) steel

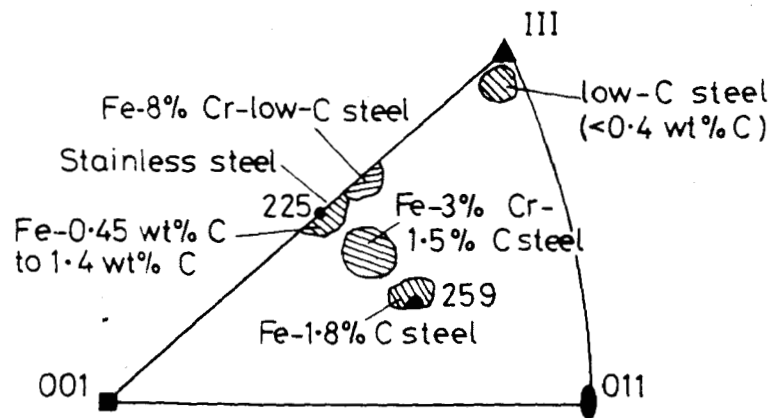


Fig. 6.11 Martensite habit planes in various types of steel.

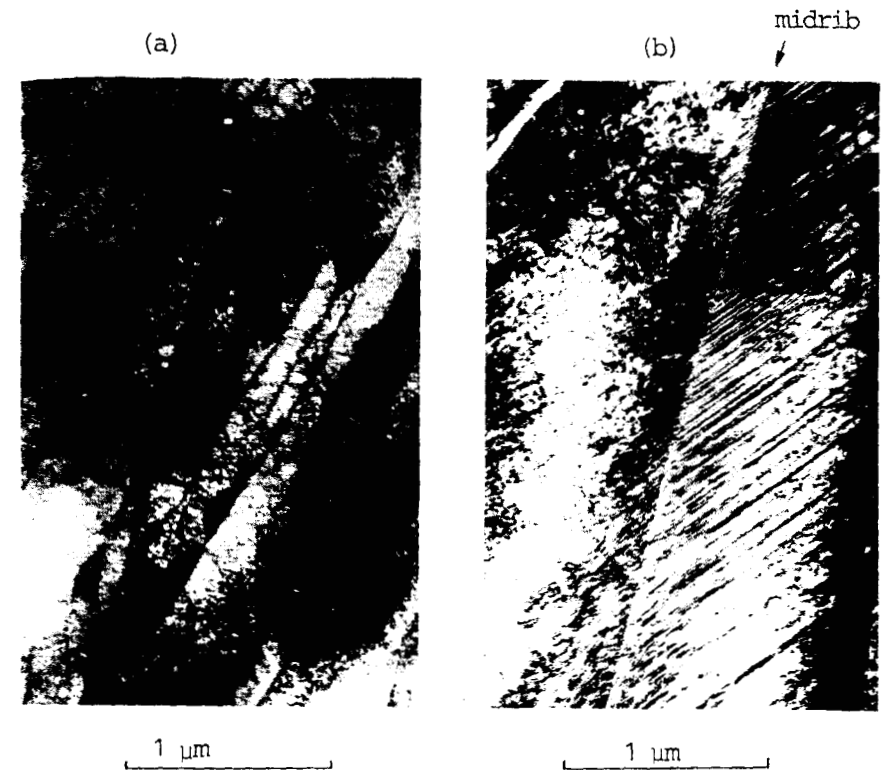


Fig. 6.12 Transmission electron micrographs of (a) lath martensite and (b) twinned martensite. Note the midrib in the twinned martensite, which is thought to be the first part of the plate to grow.

for example, to achieve a $\{259\}_\gamma$ habit plane, twins having a spacing of only 8–10 atomic planes, or ~ 3 nm, are predicted. Twin thicknesses of this order of magnitude are observed in electron micrographs of high carbon martensites. On the other hand, it is usually difficult to predict exactly the habit plane of a given alloy on the basis of known lattice parameters, dilations, etc., and apart from a few cases, the theory is mainly of qualitative interest. The theory is essentially phenomenological, and should not be used to interpret the kinetics of the transformation. Attempts at combining the crystallographical aspects of the transformation with the kinetics have, however, recently been made and will be discussed later.

6.3 Theories of Martensite Nucleation

A single plate of martensite in steel grows in 10^{-5} to 10^{-7} s to its full size, at velocities approaching the speed of sound. Using resistivity changes to monitor the growth of individual plates of martensite in, e.g. Fe–Ni alloys,

speeds of 800–1100 m/s have been measured (Nishiyama, 1978)⁶. The nucleation event is thus very important in martensitic transformations because of its likely influence on the final form of the full-grown plate. This implies that the nucleation of martensite influences the strength and toughness of martensitic steels, since for a given austenite grain size, if the number of nuclei is large, then the final grain size of the martensite will be finer and hence the steel may be stronger.

Because of the great speed of growth of martensite, it is extremely difficult to study this transformation experimentally. An example of Bunshah and Mehl's (1953) resistivity measurements⁷ is shown schematically in Fig. 6.13, indicating that α' gives a lower resistivity than γ . The small initial increase in resistivity is explained in terms of the initial strain of the austenite lattice by the martensite nucleus. This suggests in turn that the initial nucleus should be *coherent* with the parent austenite. This factor could be an important starting point when considering how nucleation occurs.

6.3.1 Formation of Coherent Nuclei of Martensite

The total increase in Gibbs free energy associated with the formation of a fully coherent inclusion of martensite in a matrix of austenite can be expressed as:

$$\Delta G = A\gamma + V\Delta G_s - V\Delta G_v \quad (6.5)$$

where γ is the interfacial free energy, ΔG_s is the strain energy, ΔG_v the volume free energy release, V the volume of the nucleus and A the surface

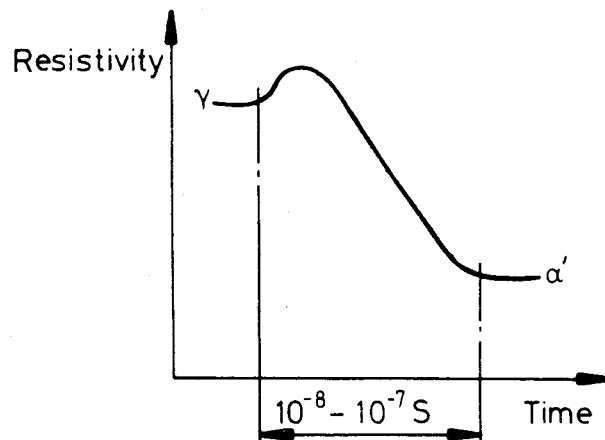


Fig. 6.13 Resistivity changes during the growth of single plates of martensite across a grain in a Fe–Ni alloy. From this it can be calculated that the velocity of growth is about 1000 m/s. (After R. Bunshah and R.F. Mehl, *Transactions AIME* 197 (1953) 1251.)

area. This expression does not account for possible additional energies that may be available due, e.g. to thermal stresses during cooling, externally applied stresses, and stresses produced ahead of rapidly growing plates. As with other nucleation events there is a balance between surface and elastic energy on the one side, and chemical (volume) free energy on the other. However, in martensitic transformations the strain energy of the coherent nucleus is much more important than the surface energy, since the shear component of the pure Bain strain is as high as $s \approx 0.32$ which produces large strains in the surrounding austenite. On the other hand, the interfacial (surface) energy of a fully coherent nucleus is relatively small.

Consider the nucleation of a thin ellipsoidal nucleus, with radius a , semi-thickness c and volume V , as illustrated in Fig. 6.14. In agreement with experimental observations, we assume that nucleation does not necessarily occur at grain boundaries. We also assume to begin with that nucleation occurs homogeneously without the aid of any other types of lattice defects. As seen from Fig. 6.14, the nucleus forms by a simple shear, s , parallel to the plane of the disc, and complete coherency is maintained at the interface. On this basis Equation 6.5 can be written

$$\Delta G = 2\pi a^2 \gamma + 2\mu V(s/2)^2 \frac{2(2-v)}{8(1-v)} \pi c/a - \frac{4}{3} \pi a^2 c \cdot \Delta G_v \quad (6.6)$$

where γ refers to the coherent interfacial energy of the coherent nucleus, v is the Poisson's ratio of the austenite, and μ is the shear modulus of the austenite. If $v = \frac{1}{3}$, Equation 6.6 can be simplified to:

$$\Delta G = \underbrace{2\pi a^2 \gamma}_{\text{surface}} + \underbrace{\frac{16\pi}{3} (s/2)^2 \mu a c^2}_{\text{elastic}} - \underbrace{\frac{4\pi}{3} a^2 c \cdot \Delta G_v}_{\text{volume}} \quad (6.7)$$

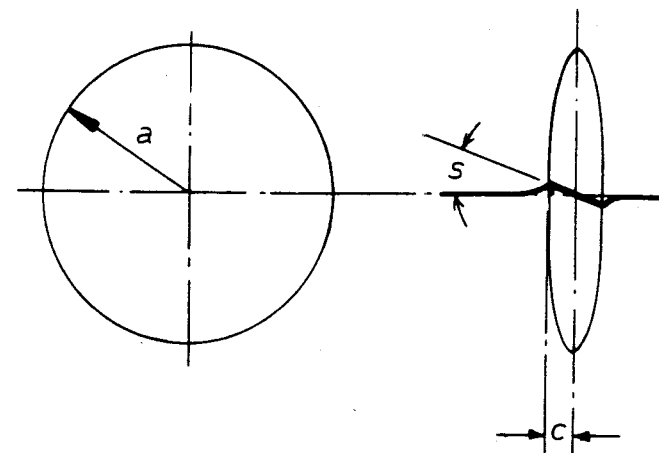


Fig. 6.14 Schematic representation of a martensite nucleus.

In this expression, the negative term ΔG_v is the free energy difference at the M_s temperature between the austenite and martensite phases and is defined in Fig. 6.3a. The middle term referring to the strain energy, is due to the shear component of strain only and neglects the small additional strain due to the dilatation which is assumed to occur normal to the disc. As pointed out by Christian (1965)¹, the most favourable nucleation path is given by the condition that the habit plane is exactly an invariant plane of the shape transformation (to reduce coherency energy) although this is not necessarily realized in practice. The minimum free energy barrier to nucleation is now found by differentiating Equation 6.7 with respect to a and c , and by subsequent substitution we obtain:

$$\Delta G^* = \frac{512}{3} \cdot \frac{\gamma^3}{(\Delta G_v)^4} \cdot (s/2)^4 \mu^2 \pi \quad \text{joules/nucleus} \quad (6.8)$$

This expression is thus the nucleation barrier to be overcome by thermal fluctuations of atoms if classical, homogeneous nucleation is assumed. It is seen that the energy barrier is extremely sensitive to the values chosen for γ , ΔG_v and s . The critical nucleus size (c^* and a^*) is also highly dependent upon these parameters. It can be shown that:

$$c^* = \frac{2\gamma}{\Delta G_v} \quad (6.9)$$

and

$$a^* = \frac{16\gamma\mu(s/2)^2}{(\Delta G_v)^2} \quad (6.10)$$

Typically $\Delta G_v = 174 \text{ MJ m}^{-3}$ for steel. s varies according to whether the net shear of a whole plate (e.g. as measured from surface markings) or the shear of a fully coherent plate (as measured from lattice fringe micrographs) is considered. For the present we shall assume a value of 0.2 which is the 'macroscopic' shear strain in steel. We can only guess at the surface energy of a fully coherent nucleus, but a value of $\sim 20 \text{ mJ m}^{-2}$ seems reasonable. Using these values gives $c^*/a^* \approx 1/40$, and $\Delta G^* \approx 20 \text{ eV}$, which in fact is too high for thermal fluctuations alone to overcome (at 700 K, $kT = 0.06 \text{ eV}$). Indeed, there is plenty of experimental evidence to show that martensite nucleation is in fact a *heterogeneous process*. Perhaps the most convincing evidence of heterogeneous nucleation is given by small particle experiments^{8,9,10}.

In these experiments small single-crystal spheres of Fe-Ni of a size range from submicron to a fraction of a millimetre were cooled to various temperatures below the M_s , and then studied metallographically. These experiments showed that:

1. Not all particles transformed even if cooled down to + 4 K, i.e. $\sim 300^\circ\text{C}$ below the M_s of the bulk material; this appears to completely rule out

homogeneous nucleation, since this should always occur at a certain undercooling. Indeed, the maximum undercooling for certain alloys reached as much as $600\text{--}700^\circ\text{C}$

2. The average number of nuclei (based on plate counts) was of the order of 10^4 per mm^3 ; this is less than to be expected for purely homogeneous nucleation.
3. The number of nuclei increases substantially with increasing supercooling prior to transformation; on the other hand, the average number of nuclei is largely independent of grain size, or even whether the particles (of a given size) are single crystals or polycrystalline.
4. The surface does not appear to be a preferred site for nucleation.

On the basis of (3) and (4) it is thought that since surfaces and grain boundaries are not significantly contributing to nucleation, then the transformation is being initiated at other defects within the crystal. The most likely types of defect which could produce the observed density of nuclei are individual dislocations, since an annealed crystal typically contains $\sim 10^5$ or more dislocations per mm^2 .

6.3.2 Role of Dislocations in Martensite Nucleation

A number of researchers have considered possible ways in which dislocations may contribute to martensite nucleation. It is instructive to consider some of these ideas and see how they can fit in with the various features of martensitic transformations already discussed.

Zener (1948)¹¹ demonstrated how the movement of $\langle 112 \rangle_\gamma$ partial dislocations during twinning could generate a thin bcc region of lattice from an fcc one, and this is illustrated in Fig. 6.15. In this figure the different layers of the close-packed planes of the fcc structure are denoted by different symbols and numbered 1, 2, 3, from bottom to top layer. As indicated, in the fcc lattice the normal twinning vector is $\bar{b}_1$, which can be formed by the dissociation of an $\frac{a}{2}\langle 110 \rangle$ dislocation into two partials:

$$\bar{b} = \bar{b}_1 + \bar{b}_2 \quad (6.11)$$

i.e.

$$\frac{a}{2}[110] = \frac{a}{6}[\bar{2}11] + \frac{a}{6}[\bar{1}2\bar{1}]$$

In order to generate the bcc structure it requires that all the 'triangular' (Level 3) atoms jumps forward by $\frac{1}{2}\bar{b}_1 = \frac{a}{12}[\bar{2}11]$. In fact, the lattice produced is not quite the bcc one after this shear, but requires an additional dilatation to bring about the correct lattice spacings. As pointed out by Christian (1965)¹, however, this reaction produces a bcc lattice only two atom layers thick. Recent electron microscopy work by Brooks *et al.* (1979)¹² indicated that

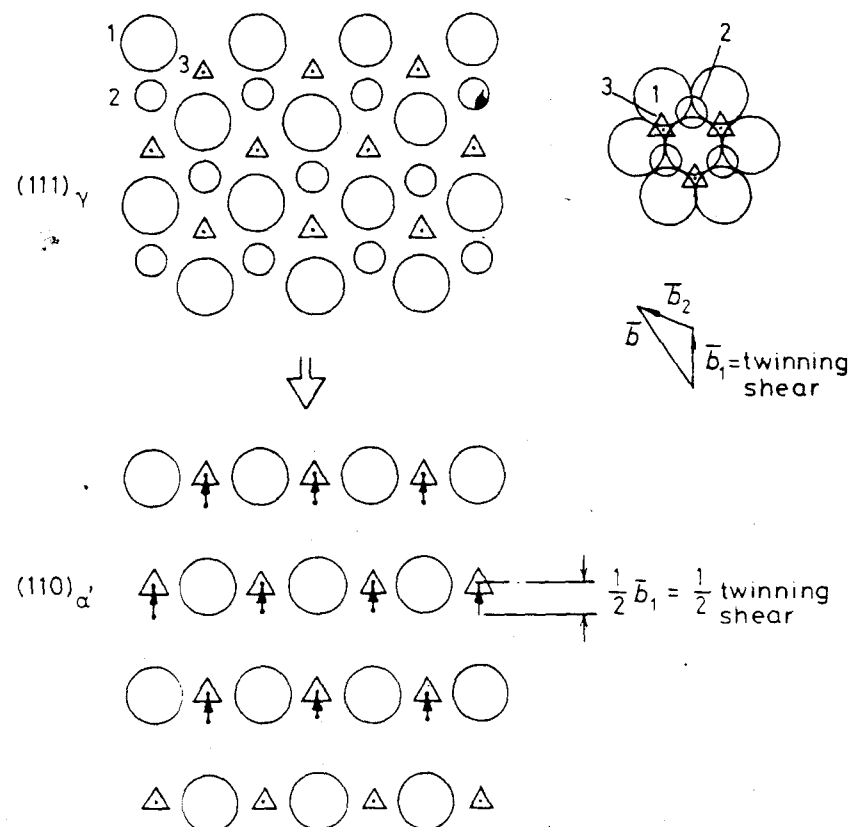


Fig. 6.15 Zener's model of the generation of two-atom-thick martensite by a half-twinning shear (some additional minor adjustments are also needed).

thicker nuclei could form by this mechanism at dislocation pile-ups, where the partial dislocations are forced closer together thereby reducing the slip vectors such that the core structures correspond to a bcc stacking. Pile-ups on nearby planes can hence interact such as to thicken the pseudo-bcc region.

An alternative suggestion was earlier made by Venables¹⁴ (1962) also in connection with the formation of martensite in stainless steels, i.e. in the case of alloys of low stacking fault energy. Venables proposed that α' forms via an intermediate (hcp) phase which he termed *epsilon martensite*, thus:



Using the same atomic symbols as before, Venables' transformation mechanism is shown in Fig. 6.16. The ϵ' -martensite structure thus thickens by inhomogeneous half-twinning shears on every other $\{111\}_\gamma$ plane. Such faulted regions have been observed to form in conjunction with martensite and an example is given in Fig. 6.17. On the other hand, there has been no direct evidence of the $\epsilon \rightarrow \alpha$ transition, and recent electron microscopy work

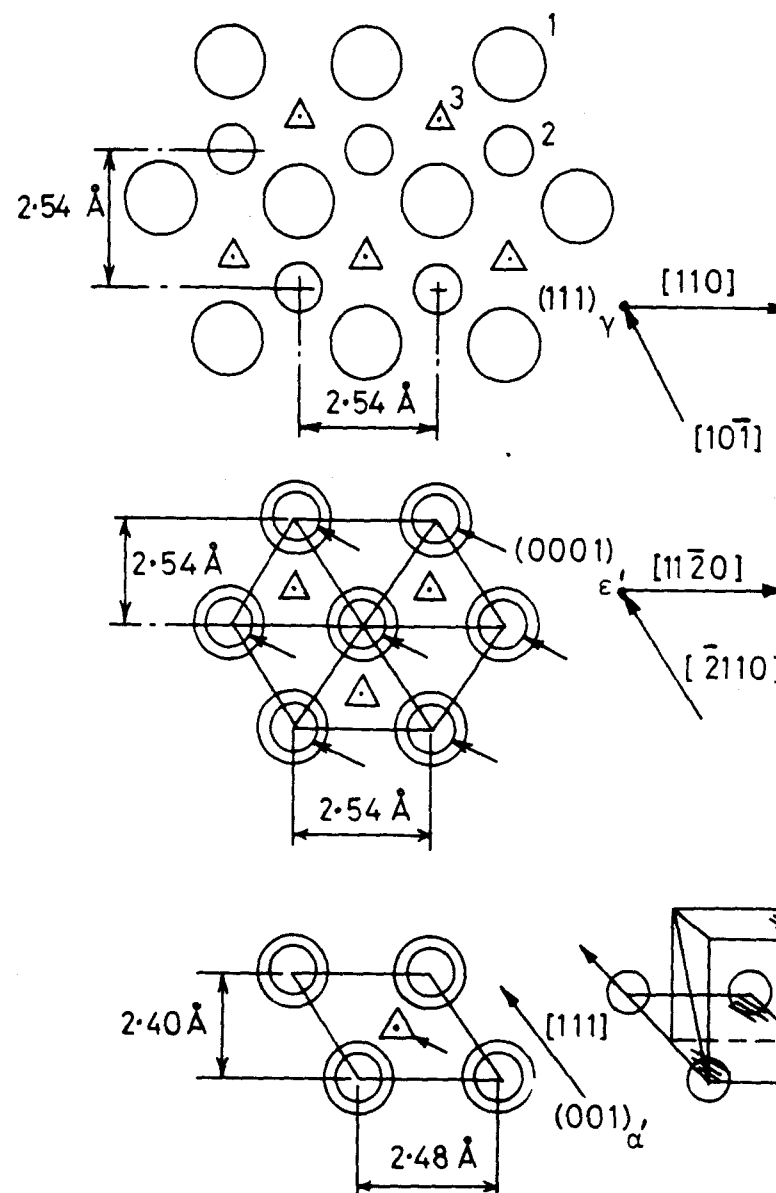


Fig. 6.16 Venables' model for the $\gamma \rightarrow \epsilon' \rightarrow \alpha'$ transformation in stainless steel.

indicates that the ϵ' and α' phases in martensitic stainless steels form independently of each other by different mechanisms, i.e. the transformation reactions in stainless steel are of the type $\gamma \rightarrow \epsilon$ or $\gamma \rightarrow \alpha'$ ^{12,14}. Other detailed models of how dislocations may bring about the martensitic transformation in iron alloys have been given, e.g. by Bogers and Burgers (1964)¹⁵ and more recently by Olson and Cohen (1976)¹⁶.



Fig. 6.17 Dislocation-assisted martensite transformation in a plastically deformed 17% Cr-8% Ni Stainless steel. (By courtesy of B. Lehtinen, Institute for Metals Research, Stockholm, Sweden.)

Another example of the fact that the half-twinning shear in fcc material can induce a martensitic transformation is in cobalt¹⁷. In this case there is an fcc $\rightarrow$ cph transformation at around 390 °C. The generation of large numbers of $\frac{1}{2}\langle 11\bar{2} \rangle_\gamma$ partial dislocations on $\{111\}_\gamma$ planes has been observed directly in the transmission electron microscope using a hot stage, as shown in Fig. 6.18. The stacking faults in this case appeared to initiate at grain boundaries. The habit plan is $\{111\}_\gamma$ and the orientation relationship is $(111)_\gamma // (0001)_\alpha$. The transformation is reversible (at ~ 430 °C) and the cph $\rightarrow$ fcc reaction occurs by the following dissociation on the hcp basal plane:



As before, the reaction has to occur on every other hcp plane in order to generate the fcc structure.

It is thus seen that some types of martensite can form directly by the systematic generation and movement of extended dislocations. It is as if the M_s temperature of these alloys marks a transition from positive to negative stacking fault energy. It appears, however, that this type of transformation

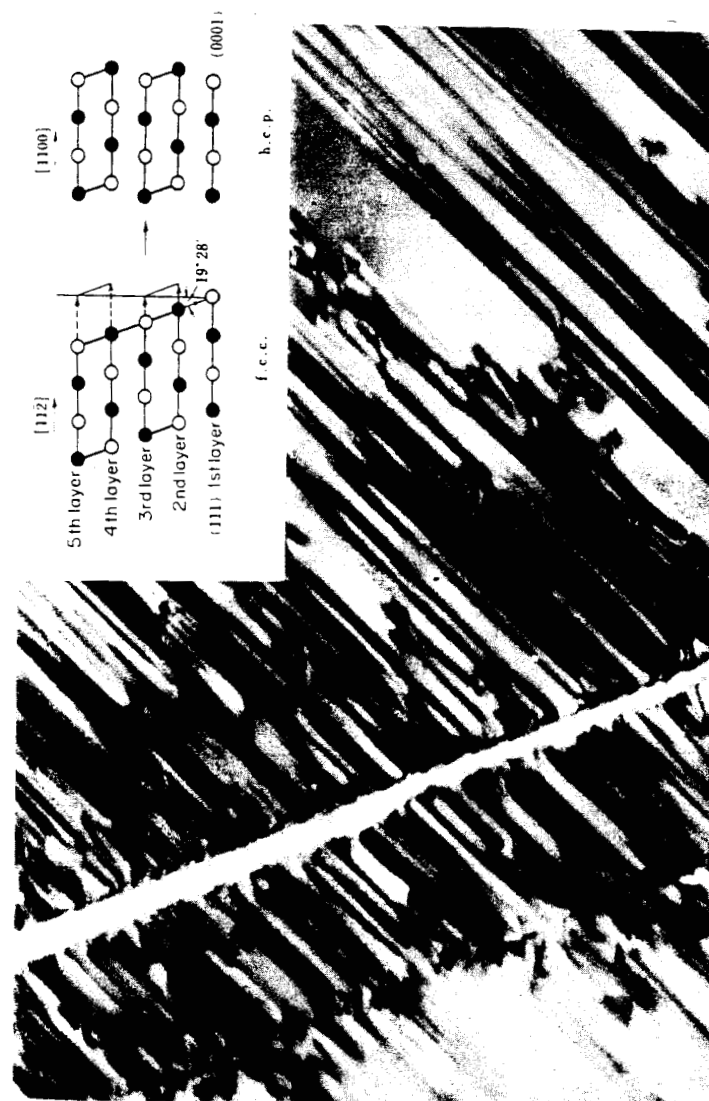


Fig. 6.18 Dislocation-assisted martensite transformation in cobalt. The insert illustrates the way stacking fault formation induces the fcc $\rightarrow$ hcp transformation. (After Z. Nishiyama, *Martensitic Transformation*, Academic Press, 1978.)

can not occur in high stacking fault energy nor in *thermoelastic martensites*, and it is thus necessary to consider alternative ways in which dislocations can nucleate martensite other than by changes at their cores. It is also difficult to understand twinned martensite, merely on the basis of dislocation core changes.

6.3.3 Dislocation Strain Energy Assisted Transformation

We now consider the possibility that the nucleation barrier to form coherent nuclei can be reduced by the help of the *elastic strain field of a dislocation*. This theory²⁰ thus differs fundamentally from the other dislocation-assisted transformation theories discussed, all of which were based on atomic shuffles within the dislocation core. We also note that in this case it is unnecessary that the habit plane of the martensite corresponds to the glide planes of austenite. Furthermore, it is assumed that coherent nuclei are generated by a pure Bain strain, as in the classical theories of nucleation.

It can be shown that the strain field associated with a dislocation can in certain cases provide a favourable interaction with the strain field of the martensite nucleus, such that one of the components of the Bain strain is neutralized thereby reducing the total energy of nucleation. This interaction is illustrated schematically in Fig. 6.19, in which it is seen that the dilatation associated with the extra half plane of the dislocation contributes to the Bain strain. Alternatively the shear component of the dislocation could be utilized.

Such an interaction thus modifies the total energy of Equation 6.5 to:

$$\Delta G = A\gamma + V\Delta G_s - V\Delta G_v - \Delta G_d \quad (6.14)$$

where ΔG_d represents the dislocation interaction energy which reduces the nucleation energy barrier. It can be shown that this interaction energy is given by the expression:

$$\Delta G_d = 2\mu s\pi \cdot ac \cdot b \quad (6.15)$$

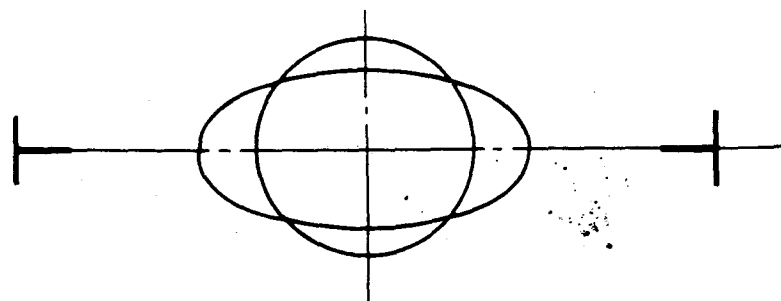


Fig. 6.19 Illustrating how one of the strain components of the Bain deformation may be compensated for by the strain field of a dislocation which in this case is tending to push atom planes together.

where $\bar{b}$ refers to the Burgers vector of the dislocation, and s refers to the shear strain of the nucleus.

The interaction energy used in Equation 6.15 assumes that a complete loop is interacting with the nucleus. In practice it is likely that only a part of a dislocation will be able to 'react' with the nucleus in this way.

Equation 6.14 may now be written in full (see Equation 6.7) as:

$$\Delta G = 2\pi a^2\gamma + \frac{16\pi}{3}(s/2)^2\mu ac^2 - \Delta G_v \frac{4\pi}{3} a^2c - 2\mu s\pi ac \cdot b \quad (6.16)$$

By summing the various components of this expression it is possible to compute the total energy of a martensite nucleus as a function of its diameter and thickness (a, c), whether it is twinned or not (this affects s , see, e.g. Fig. 6.9) and the degree of assistance from the strain field of a dislocation (or group of dislocations). This result is shown schematically in Fig. 6.20a. It has been calculated that a fully coherent nucleus can reach a size of about 20 nm diameter and two to three atoms in thickness by this partial interaction with the strain field of a dislocation. However, it will not be able to thicken or even grow larger unless twins form or slip occurs to further reduce strain energy. The attractive feature of this theory is that it essentially combines the crystallographic characteristics of the inhomogeneous shear and the Bain strain in terms of total strain energy at nucleation. It is thus in line with the majority of the known characteristics of martensite, including the initial straining of the lattice due to the coherent nucleus (see Fig. 6.13) and the fact that an inhomogeneous shear is necessary for growth. It even shows that in principle nucleation can occur in the vicinity of any dislocation, thus underlining the

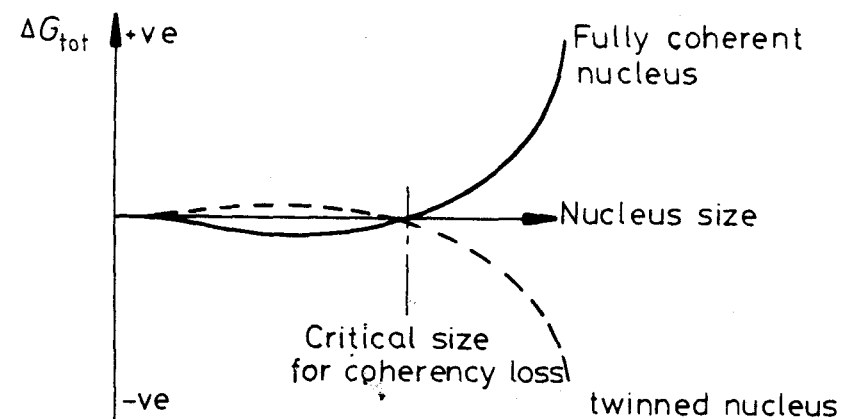


Fig. 6.20 (a) Schematic diagram based on Equation 6.16, illustrating the need for the nucleus to twin if it is to grow beyond a certain critical size. (b) Lattice image of the tip of a martensite plate in a Ti-Ni alloy. The first interfacial dislocation behind the growing front is indicated. (After R. Sinclair and H.A. Mohamed, *Acta Metallurgica* 26 (1978) 623.

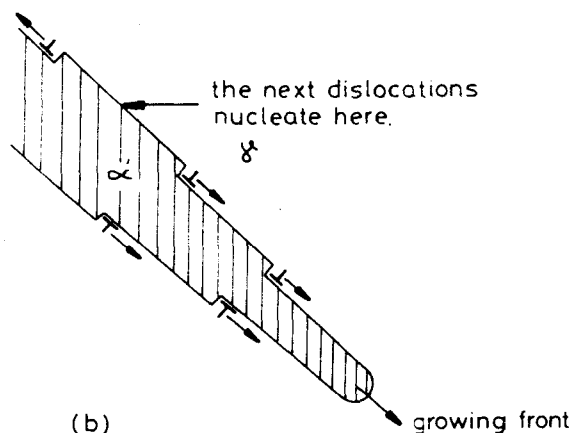
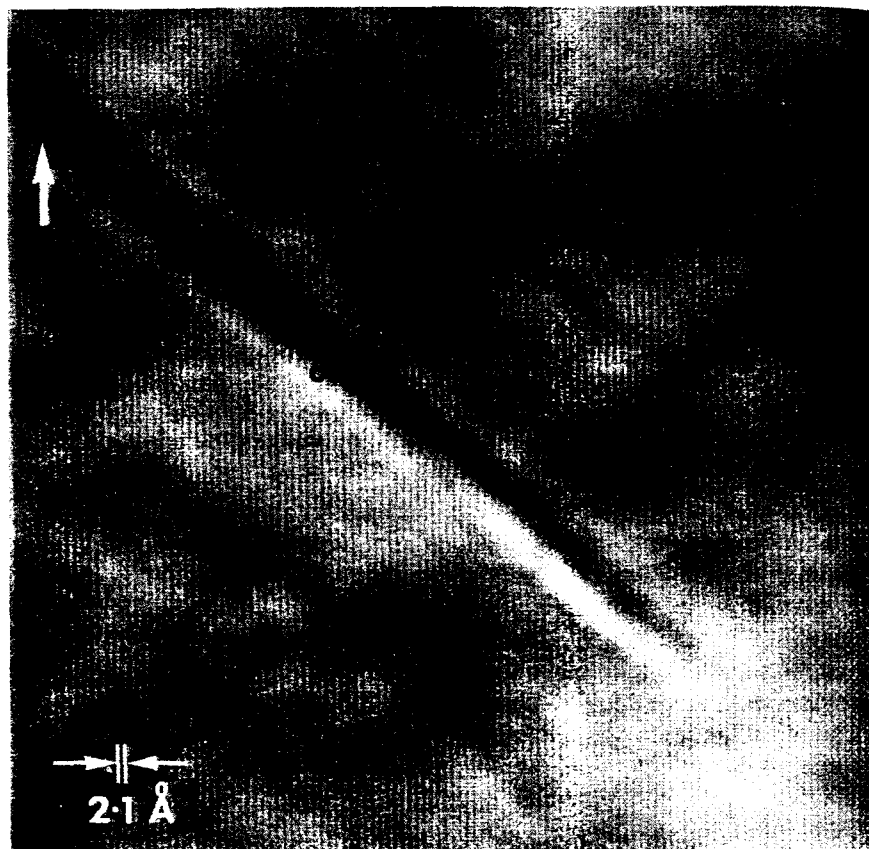


Fig. 6.20 (continued)

statistical correlation between dislocation density in well-annealed austenite and martensite formation indicated by the small particle experiments.

The M_s temperature is thus associated with the most potent nuclei, perhaps depending on the orientation or configuration of the dislocation, or groups of dislocations with respect to the potential martensite nucleus. The large undercoolings below bulk M_s as observed from the small particle experiments thus reflect the statistical probability that ideally oriented dislocations are relatively few and far between, so that high chemical driving forces are needed in most cases. The burst phenomenon, in which an autocatalytic process of rapid, successive plate formation occurs over a small temperature range in, e.g. Fe-Ni alloys, is explicable on this basis by the large elastic stresses set up ahead of a growing plate. In this case, the elastic strain field of the plate acts as the necessary interaction term in Equation 6.10. The question of whether slip or twinning occurs at the critical nucleus size in order to assist growth of the nucleus appears to be a function of the alloy content and M_s temperature, and this factor will be taken up in more detail in the next section on martensite growth.

In summary, we have not dealt with all the theories of martensite nucleation in this section as recorded in the literature, or even with all alloys exhibiting martensitic transformations. Instead we have attempted to illustrate some of the difficulties associated with explaining a complex event which occurs at such great speeds as to exclude experimental observation. A general, all-embracing theory of martensite nucleation has still evaded us, and may not even be feasible.

6.4 Martensite Growth

Once the nucleation barrier has been overcome, the chemical volume free energy term in Equation 6.10 becomes so large that the martensite plate grows rapidly until it hits a barrier such as another plate, or a high angle grain boundary. It appears from observations, that very thin plates first form with a very large a/c ratio (see Fig. 6.14) and then thicken afterwards. In high carbon martensites this often leaves a so-called "midrib" of fine twins, and an outer less well defined region consisting of fairly regular arrangements of dislocations. In low carbon lath martensite, transmission microscopy reveals a high dislocation density, sometimes arranged in cellular networks in the case of very low C content, but no twins (see Fig. 6.12). In very high carbon martensite (259 type), only twins are observed.

In view of the very high speeds of growth, it has been conjectured that the interface between austenite and martensite must be a glissile semicoherent boundary consisting of a set of parallel dislocations or twins with Burgers vector common to both phases, i.e. *transformation dislocations*. The motion of the dislocations brings about the required lattice invariant shear transformation. As noted in Section 3.4.5, the motion of this interface may or may not generate an irrational habit plane.

The habit plane transition in steels and Fe-Ni alloys as a function of alloying content of: $\{111\}_\gamma$ lath $\rightarrow \{225\}_\gamma$ mixed lath/twins $\rightarrow \{259\}_\gamma$ twinned martensite, is not properly understood. An important factor is thought to be that increased alloying lowers the M_s temperature and that it is the temperature of transformation that dictates the mode of lattice invariant shear. Qualitatively, the slip-twinning transition in a crystal at low temperatures is associated with the increased difficulty of nucleating whole dislocations needed for slip. It is thought that the critical stress needed for the nucleation of a partial twinning dislocation is not so temperature dependent as the Peierls stress for a perfect dislocation¹⁹. On the other hand, the chemical energy available for the transformation is largely independent of M_s temperature. This implies that as the M_s temperature is lowered the mechanism of transformation chosen is governed by the growth process having least energy. The other factor affecting mode of growth, as discussed in the previous section, is how the nucleus forms. If the nucleus forms by the generation of a homogeneous Bain deformation, the orientation of the nucleus in the austenite is again dependent upon it finding the lowest energy. This may not coincide with a normal glide plane in the austenite—and in highly alloyed systems it evidently does not. On the other hand, the inhomogeneous shear during growth has to be dictated by the normal modes of slip or twinning available. This suggests that if the habit plane of the martensite is irrational, it may have to grow in discrete steps which are themselves developed by conventional modes of deformation. The resulting plate would then be, for example, likened to a sheared-over pack of cards (see, e.g. Fig. 6.9b). We now consider the two main cases of rational (lath) and irrational (plate) martensite growth in steel in more detail.

6.4.1 Growth of Lath Martensite

The morphology of a lath with dimensions $a > b \gg c$ growing on a $\{111\}_\gamma$ plane (see Fig. 6.20b) suggests a thickening mechanism involving the nucleation and glide of transformation dislocations moving on discrete ledges behind the growing front. This picture of growth is suggested, e.g. in the work of Sinclair and Mohammad (1978)²⁰ studying NiTi martensite and Thomas and Rao (1978)²¹ in the case of steel martensite.

It seems possible that due to the large misfit between the bcc and fcc lattices dislocations could be self-nucleated at the lath interface. The criterion to be satisfied for dislocation nucleation in this case is that the stress at the interface exceeds the theoretical strength of the material.

It can be shown using Eshelby's approach²² that for a thin ellipsoidal plate in which $a \gg c$ the maximum shear stress at the interface between the martensite and austenite due to a shear transformation is given by the expression:

$$\sigma = 2\mu sc/a \quad (6.17)$$

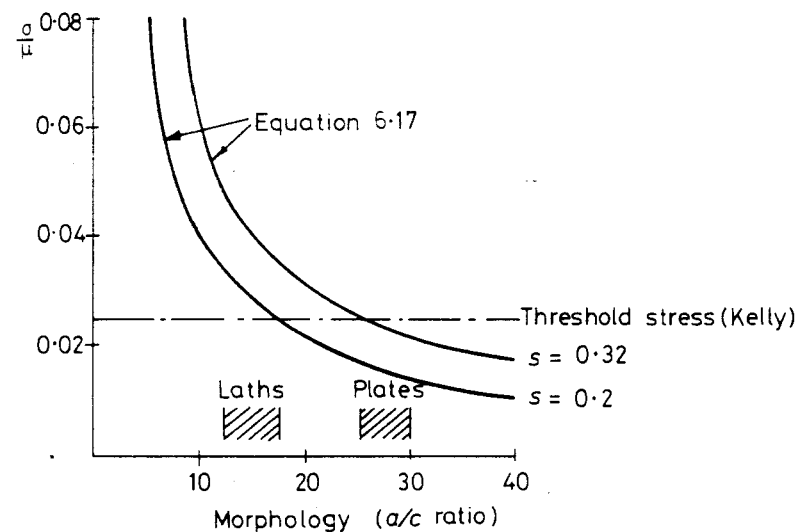


Fig. 6.21 Equation 6.17 plotted for two values of shear corresponding to a pure Bain deformation (0.32) and a twinned plate (0.2).

where μ is the shear modulus of the austenite. It is seen in this simple model that the shear stresses are sensitive to particle shape as well as angle of shear. Of course in practice it is very difficult to define the morphology of martensite in such simple c/a terms, but this gives us at least a qualitative idea of what may be involved in the growth kinetics of martensite.

Kelly (1966)²⁵ has calculated a theoretical shear strength for fcc materials of 0.025μ at ambient temperature, and this can be used as a minimum, or threshold stress for nucleating dislocations. Equation 6.17 is plotted in Fig. 6.21 in terms of different $a:c$ ratios, assuming $s = 0.2$ which is typical of bulk lath and plate martensite. An approximate range of morphologies representative of lath or plate martensite is given in the figure. It is seen that Kelly's threshold stress for dislocation nucleation may be exceeded in the case of lath martensite, but seems unlikely in the case of the thinner plate martensite. It is interesting to note from Fig. 6.21, however, that shear loop nucleation in plate martensite is feasible if $s \geq 0.32$, which is the shear associated with a pure Bain strain (Fig. 6.9a). In other words, coherency loss of the initial coherent nucleus is energetically possible.

The assumption of shear loop nucleation in fact seems reasonable and likely in conjunction with lath growth. The same mechanism of dislocation generation during growth could even be applied to bainite where the morphology appears to be fairly similar to lath martensite, although in this case some diffusion of carbon also occurs. It is thus seen that by nucleating dislocations at the highly strained interface of the laths, the misfit energy can be reduced and the lath is able to continue to grow into the austenite.

Internal friction measurements have shown that in lath martensite the density of carbon is slightly higher at cell walls than within cells, suggesting that limited diffusion of carbon takes place following or during the transformation. The transformation could also produce adiabatic heating which may affect diffusion of carbon and dislocation recovery, at least at higher M_S temperatures. In this respect there appears to be a certain relationship between lower bainite and martensite. The higher M_S temperatures associated with lath martensite may be sufficient to allow dislocation climb and cell formation after the transformation, although the high growth speeds suggest an interface of predominantly screw dislocations. The volume of retained austenite between laths is relatively small in lath martensite (these small amounts of retained austenite are now thought to be important to the mechanical properties of low-carbon steels²⁴), suggesting that sideways growth, and transformation between laths occurs without too much difficulty.

6.4.2 Plate Martensite

In medium and high carbon steels, or high nickel steels, the morphology of the martensite appears to change from a lath to a roughly plate-like product. This is associated with lower M_S temperatures and more retained austenite, as illustrated by Fig. 6.22. However, as mentioned earlier, there is also a transition from plates growing on $\{225\}_\gamma$ planes to $\{259\}_\gamma$ planes with increasing alloy content. The lower carbon or nickel $\{225\}_\gamma$ martensite often consists of plates with a central twinned 'midrib', the outer regions of the plate being free of twins. It appears that the twinned midrib forms first and the outer (dislocation) region which is less well defined than the midrib, grows afterwards. The high carbon or nickel $\{259\}$ martensite on the other hand is completely

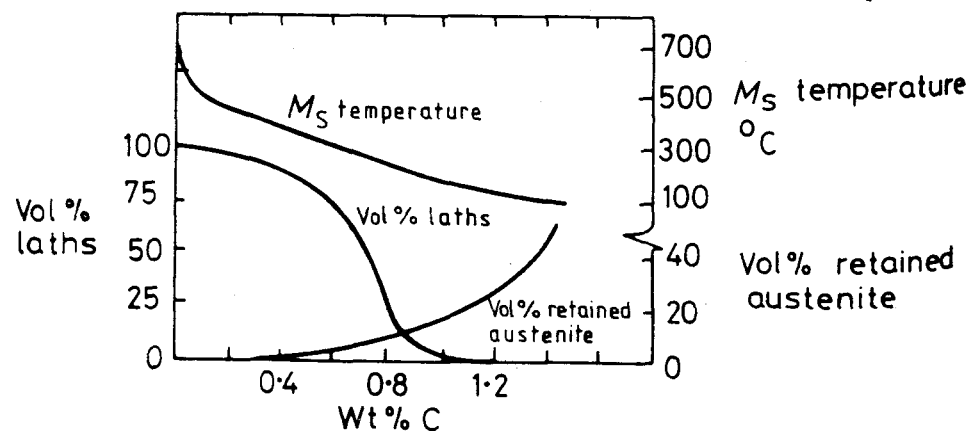


Fig. 6.22 Approximate relative percentages of lath martensite and retained austenite as function of carbon content in steels. (Data from G.R. Speich, *Metallurgical Transactions*, 3 (1972) 1045.)

twinned and the habit plane measurements have less scatter than the mixed structures.

Typical morphologies for plate martensite are usually thought to be much thinner than lath martensite or bainite. On the basis of Fig. 6.21 it appears that there is likely to be a problem in nucleating whole dislocations in the case of growing plate martensite when $s \approx 0.2$, but that partial twinning dislocations evidently are able to nucleate. Once nucleated, twinned martensite grows extremely rapidly, but the mechanism by which this occurs has not been clarified as yet. It is clear from work on low temperature deformation of fcc metals, that twinning can be an important deformation mechanism. However, the problem in martensite transformations is to explain the extremely rapid rates of plate growth as based on twinning mechanisms. The pole mechanism seems inadequate in this respect, although mechanisms based on dislocation reflection processes may be more realistic²⁷. Alternatively, it may be necessary to invoke theories in which standing elastic waves may nucleate twinning dislocations²⁵ as an aid to very rapid plate growth.

The transition from twinning $\rightarrow$ dislocations in 'midrib martensite' is intriguing and could be the result of a change in growth rate after the midrib forms (see, e.g., Shewmon, 1969)²⁶. In other words, martensite formed at higher temperatures or slower rates grows by a slip mechanism, while martensite formed at lower temperatures and higher growth rates grows by a twinning mode. Indeed, in the case of ferritic steels, the normal mode of plastic deformation is very much a function of strain rate and temperature.

An elegant model for a dislocation generated $\{225\}_\gamma$ martensite has been postulated by Frank (1953)²⁷. Frank has basically considered the way to interface the fcc austenite lattice with that of the bcc martensite such as to reduce lattice misfit to a minimum. He finds that this can be achieved quite well with the help of a set of dislocations in the interface. In this model, the close-packed planes of the fcc and bcc structures are envisaged to meet approximately along the martensite habit plane as shown in Fig. 6.23a. Since the (111) and $(101)_\alpha$ planes meet edge-on at the interface, the close-packed directions are parallel and lie in the interface plane. The reason for the rotation, ψ , shown in Fig. 6.23a, is to equalize the atomic spacings of the (111) and $(101)_\alpha$ planes at the interface. However, in spite of this, there is still a slight misfit along the $[011]_\gamma$, $[1\bar{1}1]_\alpha$ direction where the martensite lattice parameter is $\sim 2\%$ less than that of austenite. Frank therefore proposed that complete matching can be achieved by the insertion of an array of screw dislocations with a spacing of six atom planes in the interface which has the effect of matching the two lattices and thus removing the misfit in this direction. This also brings about the required lattice-invariant shear on the $(112)_\alpha$ plane as the interface advances. The resulting interface is illustrated in Fig. 6.23b.

In terms of the minimum shear stress criterion (Fig. 6.21) the further expansion and thickening of a $\{225\}_\gamma$ twinned midrib by a Frank dislocation

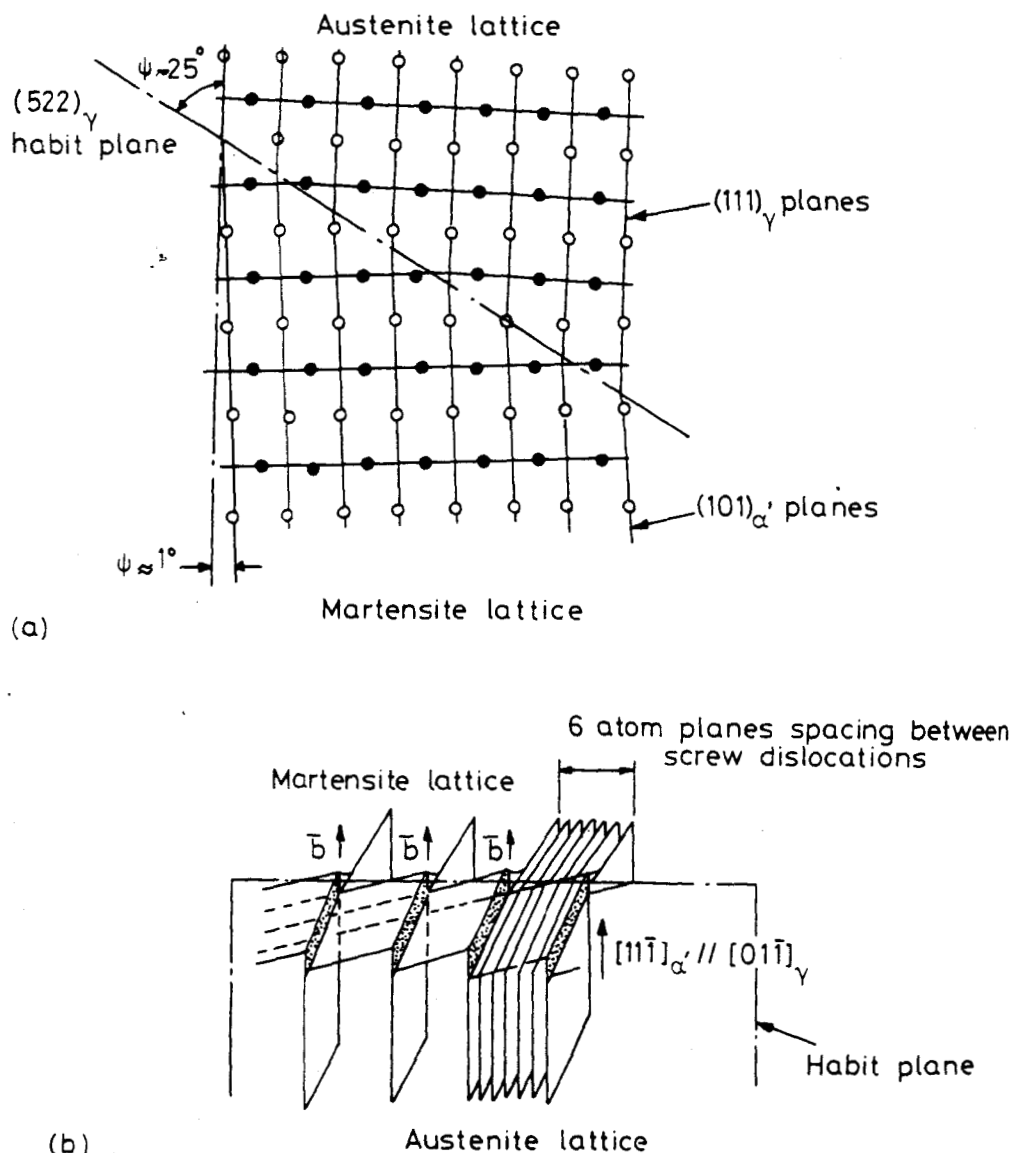


Fig. 6.23 Model for the {225} habit austenite-martensite interface in steel. (Based on data by F. C. Frank, *Acta Metallurgica* 1 (1953) 15.)

interface could occur when the midrib reaches some critical a/c ratio. However, there have been no detailed models developed as to how the Frank interface can be generated from the nucleation event. Assuming a coherent nucleus with $s \approx 0.32$, it is seen from Fig. 6.21 that it is theoretically possible for dislocation nucleation to occur at this stage to relieve coherency. Qualitatively, the larger amount of chemical free energy available after the

critical size for growth has been exceeded, may be sufficient to homogeneously nucleate dislocations²⁸ particularly in the presence of the large strain energy of the rapidly growing plate²⁹.

Other factors known to affect the growth of martensite are grain size, external stresses and the phenomenon of stabilization. We now briefly consider these effects.

6.4.3 Stabilization

This is a phenomenon associated with samples cooled to some temperature intermediate between M_s and M_f , held there for a period of time and then cooled again. In such a case, transformation does not immediately continue, and the total amount of transformed martensite is less than obtained by continuous cooling throughout the transformation range. It has even been observed that existing plates do not continue to grow after stabilization, but new plates are nucleated instead. The degree of stabilization is a function of the time held at temperature. This phenomenon is not properly understood, although it seems conceivable that carbon has time to diffuse to the interface under the influence of the high stresses associated with plate growth. There could also be local atomic relaxation at the interface, thereby increasing the nucleation barrier for dislocation generation.

6.4.4 Effect of External Stresses

In view of the dependence of martensite growth on dislocation nucleation, it is expected that an externally applied stress will aid the generation of dislocations and hence the growth of martensite. It is well established, for example, that external stress lowers the nucleation barrier for coherency loss of second phase precipitates. External stresses can also aid martensite nucleation if the external elastic strain components contribute to the Bain strain. This will provide yet another interaction term in Equation 6.14. It has been shown in such cases that the M_s temperature can be raised³⁰. However, if plastic deformation occurs, there is an upper limiting value of M_s defined as the M_d temperature. Conversely the M_s temperature can be suppressed to lower temperatures by, e.g. holding the sample being transformed under hydrostatic compression. This is because increasing pressure stabilizes the phase with the smaller atomic volume, i.e. the close-packed austenite, thereby lowering the driving force ΔG_v for the transformation to martensite. On the other hand, the presence of a large magnetic field can raise the M_s temperature on the grounds that it favours the formation of the ferromagnetic phase.

Plastic deformation of samples can aid both nucleation and growth of martensite, but too much plastic deformation may in some cases suppress the

transformation. Qualitatively it could be expected that increases in dislocation density by deformation should raise the number of potential nucleation sites, but that too much deformation may introduce restraints to nuclei growth.

The effect of plastically deforming the austenite prior to transformation on increasing the number of nucleation sites and hence refining plate size is of course the basis of the ausforming process. The high strength of ausformed steels is thus due to the combined effect of fine plate size, solution hardening (due to carbon) and dislocation hardening.

6.4.5 Role of Grain Size

Since martensite growth relies on maintaining a certain coherency with the surrounding austenite, a high-angle grain boundary is an effective barrier to plate growth. Thus while grain size does not affect the number of martensite nuclei in a given volume, the final martensite plate size is a function of the grain size. Another important feature of grain size is its effect on residual stress after transformation is completed. In large grain sized material the dilatational strain associated with the transformation causes large residual stresses to be built up between adjacent grains and this can even lead to grain-boundary rupture (quench cracking) and substantially increase the dislocation density in the martensite. Fine grain-sized metals tend to be more self-accommodating and this, together with the smaller martensite plate size, provides for stronger, tougher material.

In summary, theories of martensite nucleation and growth are far from developed to a state where they can be used in any practical way—such as helping to control the fine structure of the finished product. It does appear that nucleation is closely associated with the presence of dislocations and the process of ausforming (deforming the austenite prior to transformation) could possibly be influenced by this feature if we knew more of the mechanism of nucleation. Growth mechanisms, particularly by twinning, are still far from clarified, however.

6.5 Pre-martensite Phenomena

This is a subject that has provoked considerable attention in recent years from researchers, and is mainly concerned with ordered compounds exhibiting order $\rightarrow$ order martensitic transformations. A useful summary of this phenomenon has been given recently by Wayman (1979)³¹. The effect has been observed in the form of anomalous diffraction effects or even diffuse streaking as well as a resistivity anomaly, e.g. in TiNi alloys. In β -brass thin foils a mottled contrast has been observed giving rise to side band reflections in diffraction patterns. In CuAu alloys the phenomenon occurs in the form of a

streaming or shimmering effect in bright field images of thin foils. The latter observation, first noted by Hunt and Pashley in 1962³², has even been interpreted as possible evidence for the appearance of fluctuating strain fields due, e.g. to a Bain deformation. Recent work in Wayman's laboratory suggests, however, that while the appearance of local reordering reactions above the M_s temperature are possible, there is still little direct evidence that the observed phenomena can be related to the initial stage of the martensitic transformation. Nevertheless, the effect is an intriguing one, particularly bearing in mind the relatively low undercoolings associated with ordered alloys (see, e.g., Table 6.1), and in this respect it could be conjectured that some process is occurring which very effectively aids the transformation of these alloys.

6.6 Tempering of Ferrous Martensites

Although the diffusionless martensite transformation is fundamental to the hardening of steel, most (if not all) technological steels have to be heat treated after the transformation in order to improve toughness and in some cases even strength. Recent years have seen notable developments in these steels, achieving in some cases very high degrees of sophistication in the form of carbide dispersions and various types of substructure strengthening. For useful reviews see, e.g., Speich³³ and Honeycombe³⁴.

The martensitic transformation usually results in a ferritic phase which is highly supersaturated with carbon and any other alloying elements that remain locked into the positions they occupied in the parent austenite. On ageing, or tempering, therefore, there is a strong driving force for precipitation. As is usual with low temperature ageing the most stable precipitate, as indicated by the equilibrium phase diagram, is not the first to appear. The ageing sequence is generally $\alpha' \rightarrow \alpha + \epsilon$ -carbide or $\alpha + \text{Fe}_3\text{C}$, depending upon the tempering temperature. It is not thought that ϵ -carbide ($\text{Fe}_{2.4}\text{C}$) decomposes directly to Fe_3C , but that the transition only occurs by the ϵ -carbide first dissolving. When strong carbide-forming alloying elements such as Ti, Nb, V, Cr, W, or Mo are present the most stable precipitate can be an alloy carbide instead of cementite. See, for example, the Fe-Mo-C phase diagram in Fig. 6.24. However, these ternary additions are dissolved substitutionally in the ferrite lattice and are relatively immobile in comparison to interstitial carbon. The precipitation of these more stable carbides is therefore preceded by the formation of ϵ -carbide and Fe_3C which can occur solely by the diffusion of carbon. Alloying elements are only incorporated into the precipitate in proportion to their overall concentration in the ferrite.

The various changes that can take place during the tempering of ferrous martensites are summarized in Table 6.3. In practice heat-treatment times are limited to a few hours and the phases that appear within these time periods depend on the temperature at which tempering occurs. Therefore

Table 6.4 Carbide Precipitation Sequences

Alloy (wt%)	Precipitation sequence
Fe-C	ϵ -carbide* $\rightarrow$ Fe ₃ C ($\rightarrow$ graphite)
Fe-2 V-0.2 C	Fe ₃ C $\rightarrow$ VC or V ₄ C ₃
Fe-4 Mo-0.2 C	FeC $\rightarrow$ Mo ₂ C $\rightarrow$ M ₆ C†
Fe-6 W-0.2 C	Fe ₃ C $\rightarrow$ W ₂ C $\rightarrow$ M ₂₃ C ₆ $\rightarrow$ M ₆ C
Fe-12 Cr-0.2 C	Fe ₃ C $\rightarrow$ Cr ₇ C ₃ $\rightarrow$ Cr ₂₃ C ₆

* Does not form when C $\leq$ 0.2%.

† M stands for a mixture of the substitutional alloying elements, in this case Fe and Mo.

structures, shapes and orientation relationships for some of these precipitates are listed in Table 6.5.

On the basis of the data given in Tables 6.3 to 6.5, we note the following features:

Carbon Segregation

As a result of the large distortion caused by the carbon atoms in the martensitic lattice there is an interaction energy between carbon and the strain fields

Table 6.5 Data Concerning Carbides Precipitated During Tempering of Martensite

Carbide	Crystal structure	Shape	Orientation relationship	Temperature of formation/°C
ϵ -carbide (Fe _{2.3} C)	hcp	laths	$(10\bar{1}1)_\epsilon // (101)_\alpha$ $[0001]_\epsilon // [011]_\alpha$	100–250
cementite (Fe ₃ C)	orthorhombic	laths	$(001)_c // (211)_\alpha$ $[100]_c // [0\bar{1}1]_\alpha$	250–700
VC–V ₄ C ₃	cubic (NaCl structure)	plates	$(100)_c // (100)_\alpha$ $[011]_c // [010]_\alpha$	~550
Mo ₂ C	hcp	—	$(0001)_c // (011)_\alpha$ $[11\bar{2}0]_c // [100]_\alpha$	~550
W ₂ C	hcp	needles	as Mo ₂ C	~600
Cr ₇ C ₃	hexagonal	spheres	—	~550
Cr ₂₃ C ₆ (M ₂₃ C ₆)	cubic	plates	$(100)_c // (100)_\alpha$ $[010]_c // [010]_\alpha$	~700
M ₆ C (Fe ₃ Mo ₃ C, Fe ₃ W ₃ C)	cubic	—	—	~700

around dislocations. In lath martensite, for example, carbon tends to diffuse to sites close to dislocations in order to lower its chemical potential. In plate martensite, however, the martensite is internally twinned and there are relatively few dislocations. In this case carbon-rich clusters or zones tend to form instead. In low-carbon low-alloy steels, martensite starts to form at relatively high temperatures and there can be sufficient time during the quench for carbon to segregate or even precipitate as ϵ -carbide or cementite.

ϵ -Carbide

The reason for the 0.2% C limit (Table 6.4) is thought to be due to the fact that the M_s temperatures of very low-carbon martensites are high enough to allow considerable carbon diffusion to lath boundaries during cooling (see, e.g., Fig. 6.3c). There is thus no carbon left in solution to precipitate out on reheating. ϵ -Carbide has a hexagonal crystal structure and precipitates in the form of laths with an orientation relationship as shown in Table 6.5, (see Fig. 6.25). This orientation relationship provides good matching between the $(101)_\alpha$ and $(10\bar{1}1)_\epsilon$ planes.

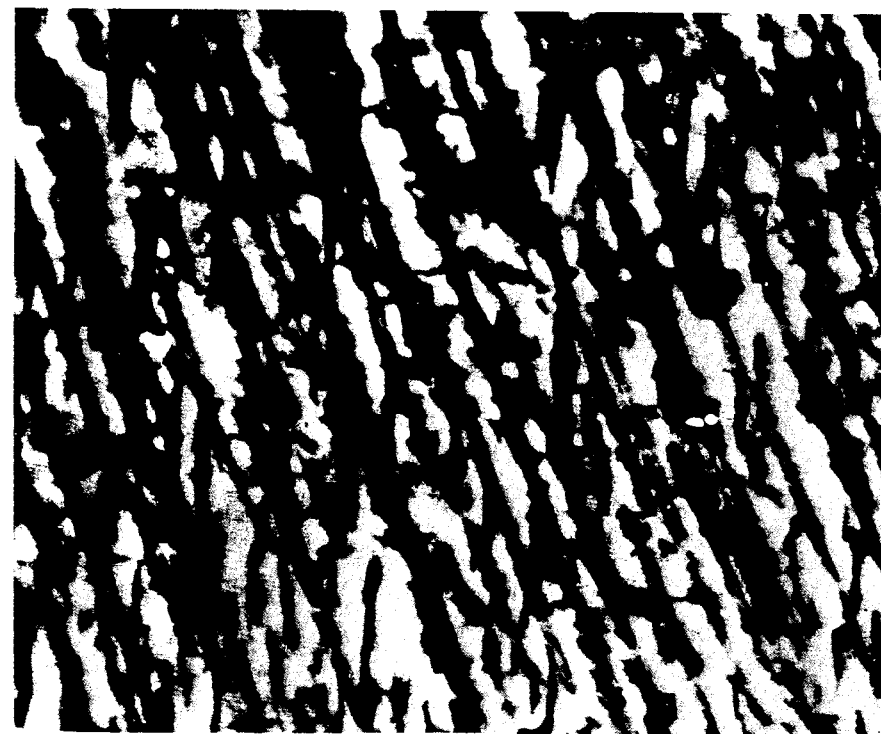


Fig. 6.25 ϵ -carbide (dark) precipitated from martensite in Fe-24 Ni-0.5 C after 30 min at 250 °C. Thin foil electron micrograph ($\times 90\,000$). (After G.R. Speich in *Metals Handbook*, 8th edn. Vol. 8, American Society for Metals, 1973, p. 202.)

Cementite

Cementite forms in most carbon steels on tempering between 250 and 700 °C. The precipitate is initially lath-like with a $\{011\}_{\alpha'}$ habit plane, Fig. 6.26. It has an orthorhombic crystal structure and forms with the orientation relationship given in Table 6.5. At high temperatures cementite rapidly coarsens into a spheroidal form, as shown in Fig. 6.32. In alloy steels the cementite composition can often be represented as $(\text{Fe}M)_3\text{C}$ where M is a carbide-forming alloying element. The composition may however be metastable if sufficient alloying elements are present.

Alloy Carbides

In steels containing sufficient carbide-forming elements alloy carbides are formed above ~500 °C where substitutional diffusion becomes significant. These carbides replace the less stable cementite which dissolves as a finer alloy carbide dispersion forms. Some typical precipitation sequences are listed in Table 6.4. There are two ways in which the $\text{Fe}_3\text{C} \rightarrow$ alloy carbide

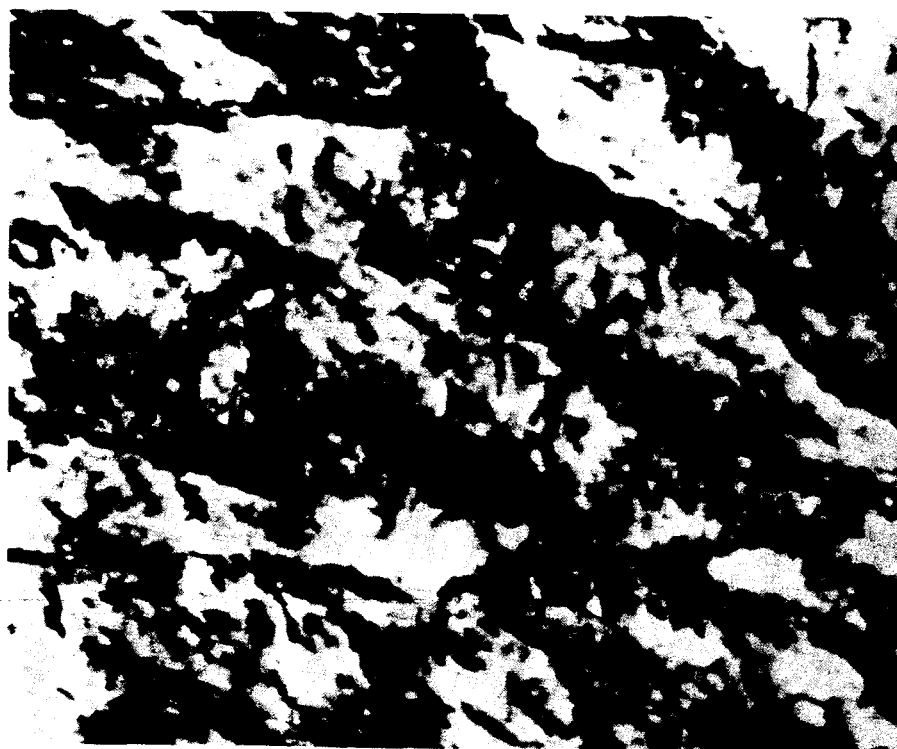


Fig. 6.26. Cementite (dark laths) formed during tempering a 0.42 C steel 1 h at 300 °C. Thin foil electron micrograph ($\times 39\,000$). (After G.R. Speich in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 202.)

transformation can take place:

1. By *in situ* transformation—the alloy carbides nucleate at several points at the cementite/ferrite interfaces, and grow until the cementite disappears and is replaced by a finer alloy carbide dispersion, see, e.g., Fig. 6.27.
2. By separate nucleation and growth—the alloy carbides nucleate heterogeneously within the ferrite on dislocations, lath boundaries, and prior austenite grain boundaries. The carbides then grow at the expense of cementite.

Either or both mechanisms can operate depending on the alloy composition.

The formation of alloy carbides is an important strengthening mechanism in high-speed tool steels that must operate at dull red heat without losing their cutting ability. The phenomenon is usually referred to as secondary hardening. Figure 6.28 shows the effect of tempering molybdenum steels for various times and temperatures. The hardness of plain carbon martensites usually decreases with increasing temperature due to recovery and overageing effects. The replacement of a coarse cementite dispersion by a finer alloy carbide that is more resistant to coarsening, however, is able to produce an increase in hardness at around 550–600 °C.

The effectiveness of these carbides as strengtheners depends on the fineness of the dispersion and the volume fraction precipitated. The fineness of the dispersion depends on ΔG^* for nucleation which in turn is influenced by the

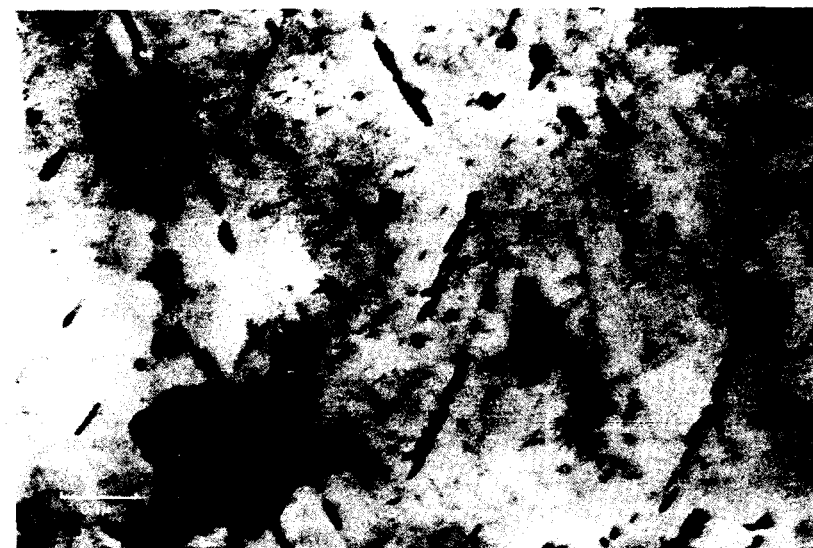


Fig. 6.27. W_2C needles lying along the sites of former Fe_3C precipitates in Fe-6.3 W-0.23 C quenched and tempered 20 h at 600 °C. (After R.W.K. Honeycombe, *Structure and Strength of Alloy Steels*, Climax Molybdenum, London, after A.T. Davenport.)

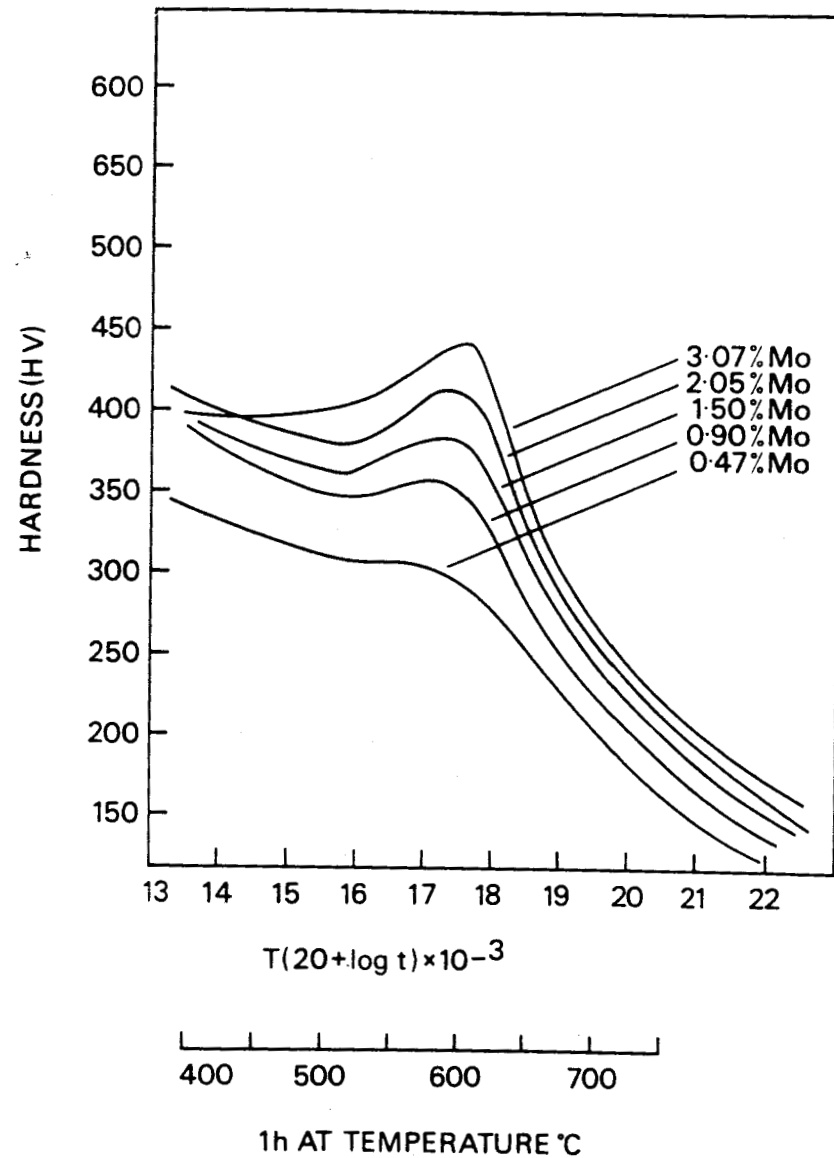


Fig. 6.28 The effect of molybdenum on the tempering of quenched 0.1% C steels. (After K.J. Irvine F.B. Pickering, *Journal of the Iron and Steel Institute* **194** (1960) 137.)

free energy of formation of the carbide, the interfacial energy and the misfit. A guide to the relative free energies of formation is given by Fig. 6.29 which shows the heats of formation (ΔH_f) of various nitrides, carbides and borides relative to that of cementite which is taken as $\Delta H_f = 0$. The finest precipitate dispersions are generally obtained from VC, NbC, TiC, TaC and HfC. These

Enthalpy of formation at 298.15 K $\Delta H_f/\text{KJ mol}^{-1}$

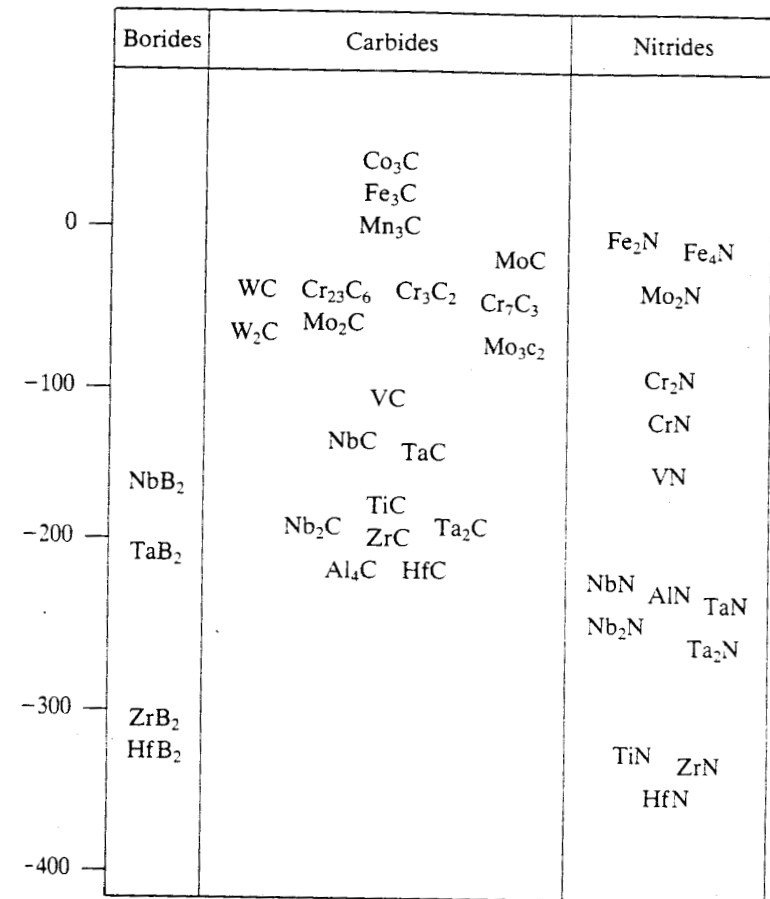


Fig. 6.29 Enthalpies of formation of carbides, nitrides and borides. (Data from H. L. Schick, *Thermodynamics of Certain Refractory Compounds*, Academic Press, 1966.)

are all close-packed intermetallic compounds. On the other hand the carbides with complex crystal structures and low heats of formation, e.g. M_7C_3 , M_6C and Mn_3C_6 , generally form relatively coarse dispersions.

The volume fraction of carbide precipitated depends on the solubility of the alloy carbide in the austenite prior to quenching, relative to the solubility in ferrite at the tempering temperature. Note that the solubility of a phase β in a terminal solid solution of α was considered in Chapter 1 for binary alloys. It can similarly be shown that in ternary Fe-C-M alloys the concentrations of M and C in Fe in equilibrium with a carbide M_mC_n are approximately given by the relation³⁸:

$$[M]^m[C]^n = K \quad (6.18)$$

where $[M]$ and $[C]$ are the atomic percentages or mole fractions of M and C in solution and K is the *solubility product* which can be expressed as

$$K = K_0 \exp \frac{-\Delta H}{RT} \quad (6.19)$$

where K_0 is a constant and ΔH is the enthalpy of formation of M_mC_n from M and C in solution.

Figure 6.30 shows the solubility products of various carbides and nitrides in austenite as a function of temperature. The solubilities of these compounds in ferrite are very much lower and to a first approximation can be considered to be approximately equal. It is clear therefore that chromium, molybdenum and vanadium with the highest solubilities in austenite, should precipitate in the highest volume fractions in the ferrite.

Effect of Retained Austenite

In most steels, especially those containing more than 0.4% C, austenite is retained after quenching. On ageing in the range 200–300 °C this austenite decomposes to bainite. In some high-alloy steels austenite can be stabilized to such low temperatures that the martensite partially reverts into austenite on heating. Very thin regions of retained austenite may even be present between laths in low-carbon steel, and this is thought to improve the toughness of these steels independently of tempering treatments.

Recovery, Recrystallization and Grain Growth

As-quenched lath martensite contains high-angle lath boundaries, low-angle cell boundaries within the laths, and dislocation tangles within the cells. Recovery usually occurs above 400 °C and leads to the elimination of both the dislocation tangles and the cell walls. The lath-like structure, however, remains as shown in Fig. 6.31. The ferrite can recrystallize at high temperatures in low-carbon steels, (see, e.g., Fig. 6.32) but the process is inhibited in medium to high-carbon steels by the grain boundary pinning caused by carbide precipitates. In the latter steels recovery is followed by grain growth.

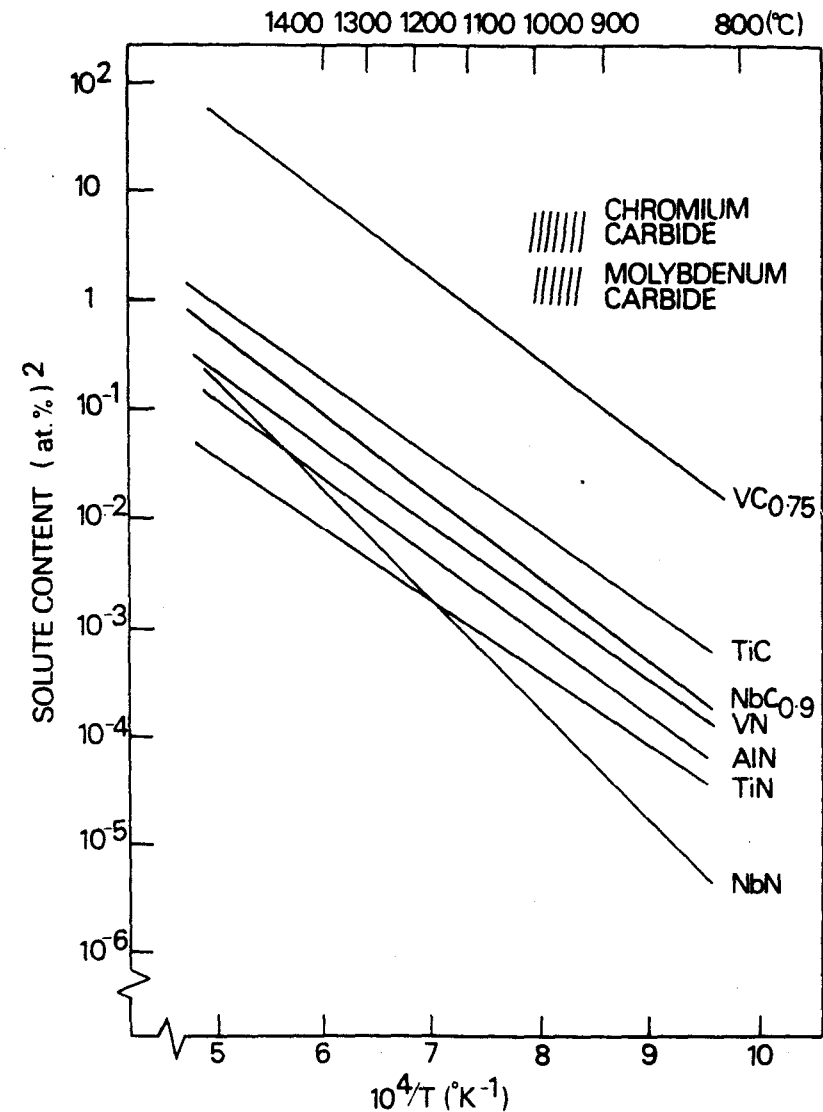


Fig. 6.30 Solubility products (in atomic per cent) of carbides and nitrides in austenite as a function of temperature. (After R.W.K. Honeycombe, *Structure and Strength of Alloy Steels*, Climax Molybdenum, London, 1973.)

Temper Embrittlement

As pointed out in the introduction to this section the aim of tempering martensite is to improve ductility. However in some steels tempering in, or slow cooling through, the range 350–575 °C can lead to embrittlement. This has been attributed to the segregation of impurity atoms such as P, Sb or Sn to prior austenite grain boundaries. Some steels also show an embrittlement on

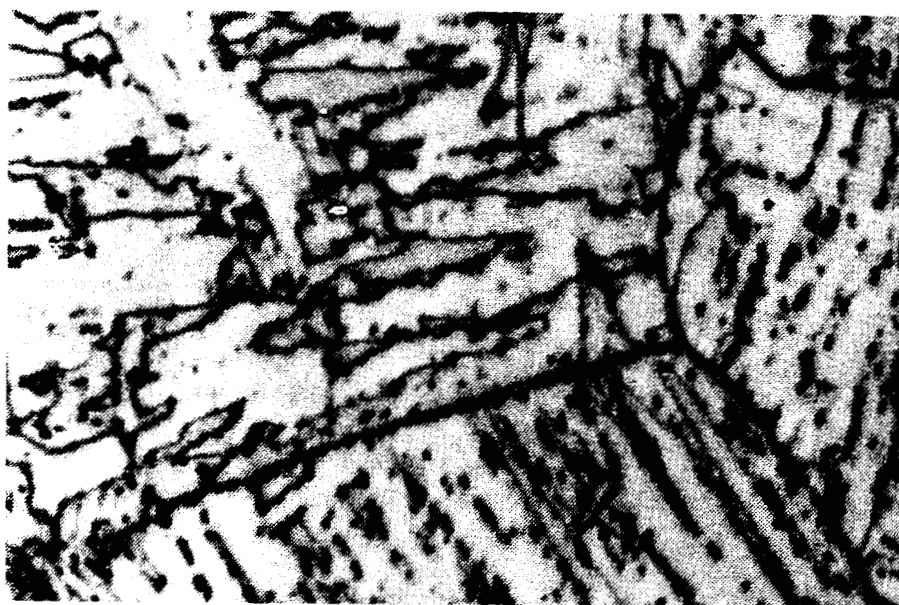


Fig. 6.31 A recovered lath martensite showing that the lath boundaries are retained. Fe-0.18 C tempered 10 min at 600 °C ($\times 2000$). (After G.R. Speich in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, Metals Park, Ohio, 1973, p. 202.)

tempering between 230 and 370 °C. This may be caused by the formation of carbides with a critical plate-like shape.

6.7 Case Studies

It is clear from the foregoing theory of martensite that much work remains to be done before we can fully understand this complex transformation, particularly in steels. In spite of this, the hardening of steel by quenching to obtain martensite is arguably one of the most important of all technological processes. In this section we illustrate four examples of *technological alloys* based on the martensite transformation. These are a quenched and tempered structural steel, some controlled transformation steels including TRIP steels, dual-phase steel, and a TiNi 'memory' metal possessing a unique shape-memory property based on a diffusionless transformation.

6.7.1 Carbon and Low-Alloy Quenched and Tempered Steels

Composition range: 0.1–0.5 wt% C; (C < 0.3%: weldable without preheat)
0.6–1.3% Mn with or without small alloying additions,
e.g. Si, Ti, Mo, V, Nb, Cr, Ni, W, etc.



Fig. 6.32 A partially recrystallized structure. Top left: recovered but not recrystallized, bottom left: a new recrystallized grain. Coarse spheroidal cementite is also apparent. Fe-0.18 C quenched and tempered 96 h at 600 °C ($\times 2000$). (After G.R. Speich, in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 202.)

Special properties: High strength, weldable constructional steels.

Relevant phase diagrams: Fe-C, in conjunction with appropriate CCT, TTT diagrams (see, e.g., Fig. 6.3).

Microstructures: See Figs. 6.1 and 6.12.

Comments: The compositions of these steels are chosen with respect to (a) hardenability; (b) weldability; (c) tempering properties, e.g. resistance to tempering, or increased tempering strength due to secondary hardening. Typically, lath or mixed (lath plus twinned) structures contain high densities of dislocations ($0.3\text{--}0.9 \times 10^{10} \text{ mm}^{-2}$), equivalent to a heavily worked steel. There is normally very little retained austenite associated with these steels see, e.g., Fig. 6.22. The lattice structure is bcc, at least for carbon contents greater than $\sim 0.2\%$. Below this composition it is suspected that due to the higher M_s temperature, some carbon segregates to dislocations or lath boundaries during the quench, as measured by resistivity and internal frictional measurements.

These results indicate that only in steels containing more than 0.2% C is carbon retained in solution. Curiously, this effect is not reflected by hardness changes and therefore the main strengthening mechanism in these steels is thought to be the fine lath or cellular structures and not so much due to

carbon in solution. The yield strength can therefore be represented by a modified version of the Hall–Petch equation (Gladman *et al.*, 1975)³⁵:

$$\sigma_y = \sigma_i + Kd^{-1/2} + \sigma_{\text{disl}} (+ \sigma_{\text{ppt}}) \quad (6.20)$$

where σ_i is the friction stress (due to alloying elements in solution) and d refers to the mean cell, or lath size. K is a constant and σ_{disl} refers to the hardening contribution due to dislocations and/or twins; σ_{ppt} refers to carbide precipitation after tempering. Typical yield strengths of these tempered steels are in the range 500–700 MN m⁻², for a mean lath width of 2–3 μm .

6.7.2 Controlled-Transformation Steels

Compositions ranges:

0.05–0.3 wt% C
0.5–2.0% Mn
0.2–0.4% Si
14.0–17.0% Cr
3.0–7.0% Ni
~2% Mo

Other possible additions: V; Cu, Co, Al, Ti, etc.

Special properties: Very high strength, weldable, good corrosion resistance; used, e.g., as skin for high speed aircraft and missiles.

Relevant phase diagrams: See Fig. 6.33.

Microstructures: Fine lath martensite with possible fine network of δ -ferrite.

Comments: Since it is required to form, or work this material at ambient temperatures prior to hardening and tempering, elements that stabilize the austenite are used in significant amounts, e.g. Ni, Cu, etc. On the other hand, the M_s – M_f range should not be depressed too far, and the relative effects of alloying elements on M_s temperature are shown in Table 6.6. It is seen that in practice very strict control over composition of these steels must be made, balancing the amount of ferrite formers (e.g. Mn) with C content. Such amounts of δ -ferrite are sometimes retained in order to improve weldability

Table 6.6 Effect of Alloying Elements on M_s in Steels

F.B. Pickering, 'Physical metallurgy of stainless steel developments' *Int. Met. Rev.*, **21**, pp. 227–268, 1976.

Element	N	C	Ni	Co	Cu	Mn	W	Si	Mo	Cr	V	* Al
Change in M_s , °C per wt%	-450	-450	-20	+10	-35	-30	-36	-50	-45	-20	-46	-53

(see Section 4.6.3). This also requires a careful balance in analyses, using, e.g., modified Schaeffler diagrams as a guide. Cold working is either carried out in the purely austenitic range (i.e. above M_d) the steel then being quenched to obtain martensite, or worked below M_d in which case deformation induces the transformation to occur without the need for refrigeration. The latter steels are known as 'transformation induced plasticity' (TRIP) steels. Since the M_s – M_f range is about 100–140 °C, the M_s temperature should not lie too far below the working temperature, or refrigeration will have to be carried out at such low temperatures that it may become too expensive. Retained austenite is undesirable in these steels because of its adverse effect on strength. The fine martensitic structure, in combination with work hardening and tempering give these steels strengths up to ~1500 MN m⁻². In Table 6.7, typical properties of various controlled transformation steels are shown as a function of the type of heat treatment and transformation.

The mechanical properties given in Table 6.7 show that samples transformed by refrigeration generally give the higher strengths. It is also seen that the austenitizing temperatures may change from alloy to alloy. Choice of austenitizing temperature is critical with regard to solution treatment, resolution of carbides and M_s temperature. For example, the lower the solution temperature, the more $M_{23}C_6$ will remain during austenitizing; this in turn reduces the Cr and C content of the austenite which raises the M_s temperature. The example given of a TRIP steel in Table 6.7 shows that this material has exceptional high strength and toughness (50% elongation).

6.7.3 The 'Shape-Memory' Metal: Nitinol

Composition range: 55–55.5 wt% Ni–44.5–45% Ti. Possible additions: small amounts of Co (to vary M_s).

Phase diagram: See Fig. 6.34.

Phase transitions:

Ordered TiNi(I) bcc A2 structure
↓ {650–700 °C diffusion controlled
TiNi (II): complex CsCl-type structure
↓ {170 °C martensitic
TiNi (III): complex structure

Special properties: The TiNi (II $\rightleftharpoons$ III) transformation is reversible and effectively enables the alloy to be deformed by a shear transformation, i.e. without irreversible plastic deformation occurring, by up to 16% elongation/contraction. Thus 'forming' operations can be made below M_s which may be 'unformed' simply by re-heating to above the M_s . These unique properties are used in such applications as,

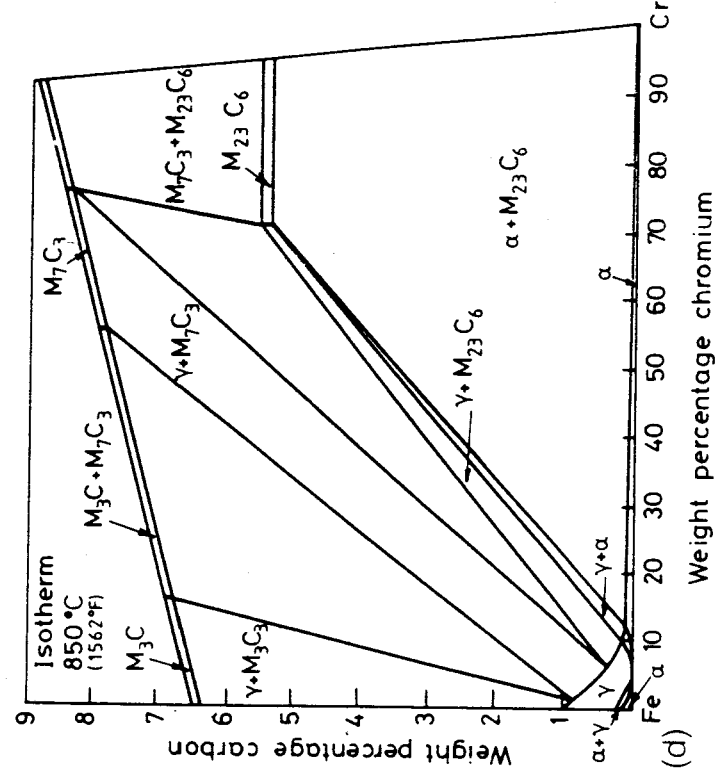
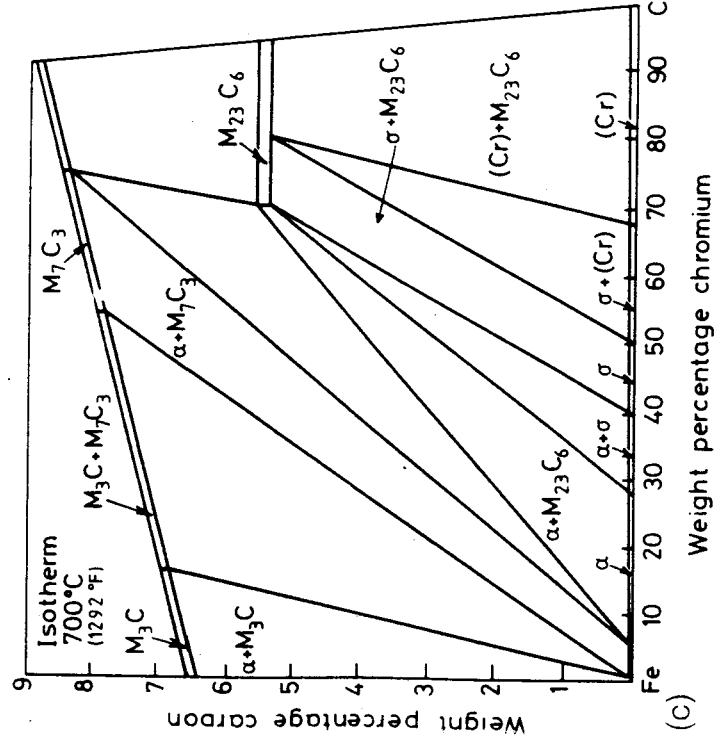
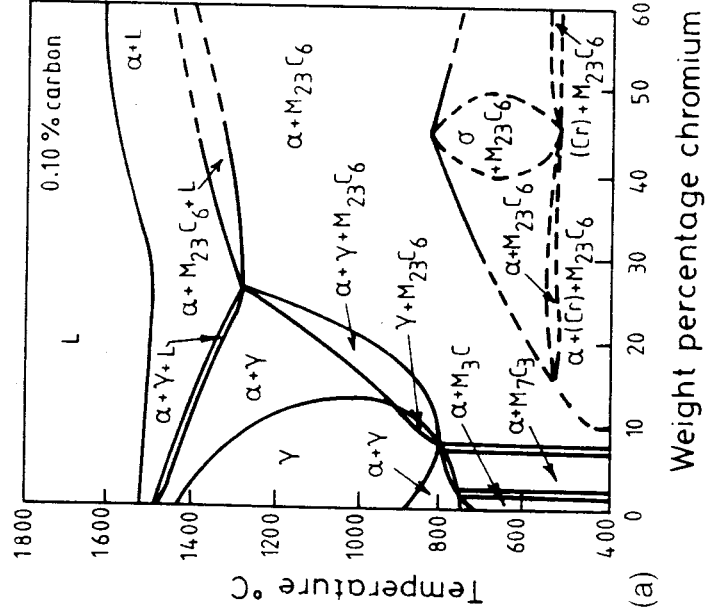
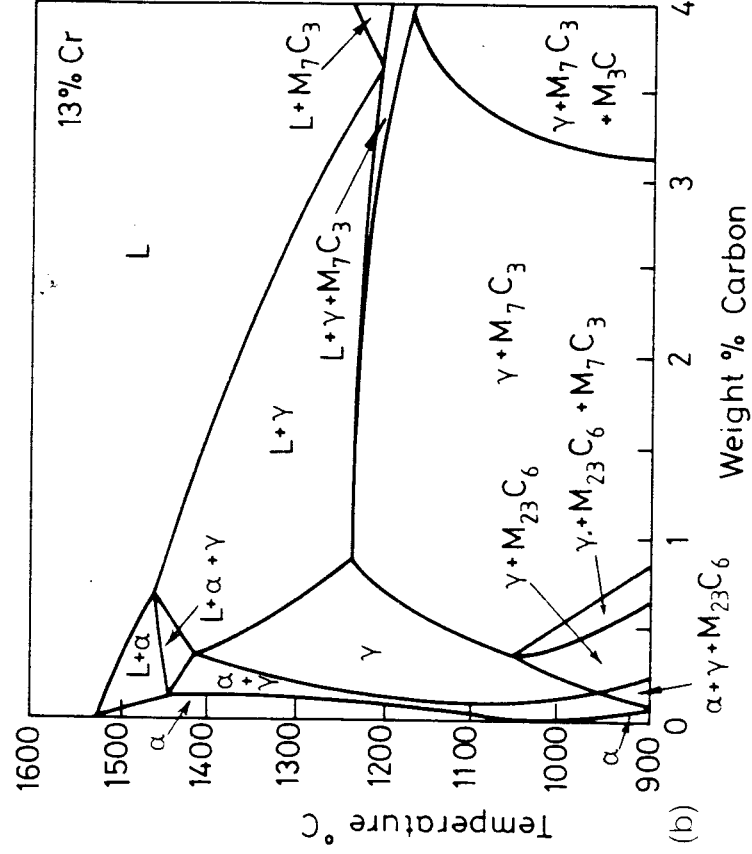


Fig. 6.33 Phase diagram for Fe-Cr-C steels ((a), (c) and (d) after W.D. Forgeng in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 402; (b) from *A Guide to the Solidification of Steels*, Jernkontoret, Stockholm, 1977.)

Table 6.7 Typical Properties of Controlled Transformation Steels

After F.B. Pickering, 'Physical metallurgy of stainless steel developments', *Int. Met. Rev.*, **21**, p. 227 (1976)

Composition	Heat treatment	Tensile strength, MN/m ²	0.2% proof stress, MN/m ²	Elongation, %
0.1C-17Cr-4Ni-3Mn (showing effect of cold working)	925 °C, tempered at 450 °C for 3 h	950	620	10
	925 °C, cold worked, 20% reduction, tempered at 450 °C for 3 h	1340	1300	5
	925 °C, cold worked, 40% reduction, tempered at 450 °C for 3 h	1700	1670	3.5
0.1C-17Cr-4Ni-3Mn	950 °C, 700 °C for 2 h, tempered at 450 °C for 2 h	1280	1110	15
	950 °C, -78 °C, tempered at 400 °C for 1 h	1440	1200	19
0.1C-17.5Cr-3Ni-3Mn-2Cu	1050 °C, 700 °C for 2 h, tempered at 400 °C for 24 h	1120	940	5
	975 °C, -78 °C, tempered at 450 °C for 24 h	1500	1260	11
0.1C-17.5Cr-4Ni-3Mn-1Al	950 °C, -78 °C, tempered at 450 °C for 4 h	1450	1260	13
	1000 °C, -78 °C, tempered at 450 °C for 4 h	1340	1150	23
0.06C-16.5Cr-2Mn-5Ni-1.5Mo-2Co-1Al	1050 °C, 700 °C for 2 h, tempered at 450 °C for 4 h	1430	1270	3
	950 °C, -78 °C, tempered at 450 °C for 4 h	1520	1240	21
0.07C-17.5Cr-3Ni-2Mn-2Mo-2Co-1Cu	1050 °C, 700 °C for 2 h, tempered at 450 °C for 4 h	1250	1110	10
	950 °C, -78 °C, tempered at 450 °C for 4 h	1360	1240	20
TRIP STEEL: 0.3C-2Mn-2Si-9Cr-8.5Ni-4.0Mo	Austenitized 425 °C, 80% reduction	1500	1430	50

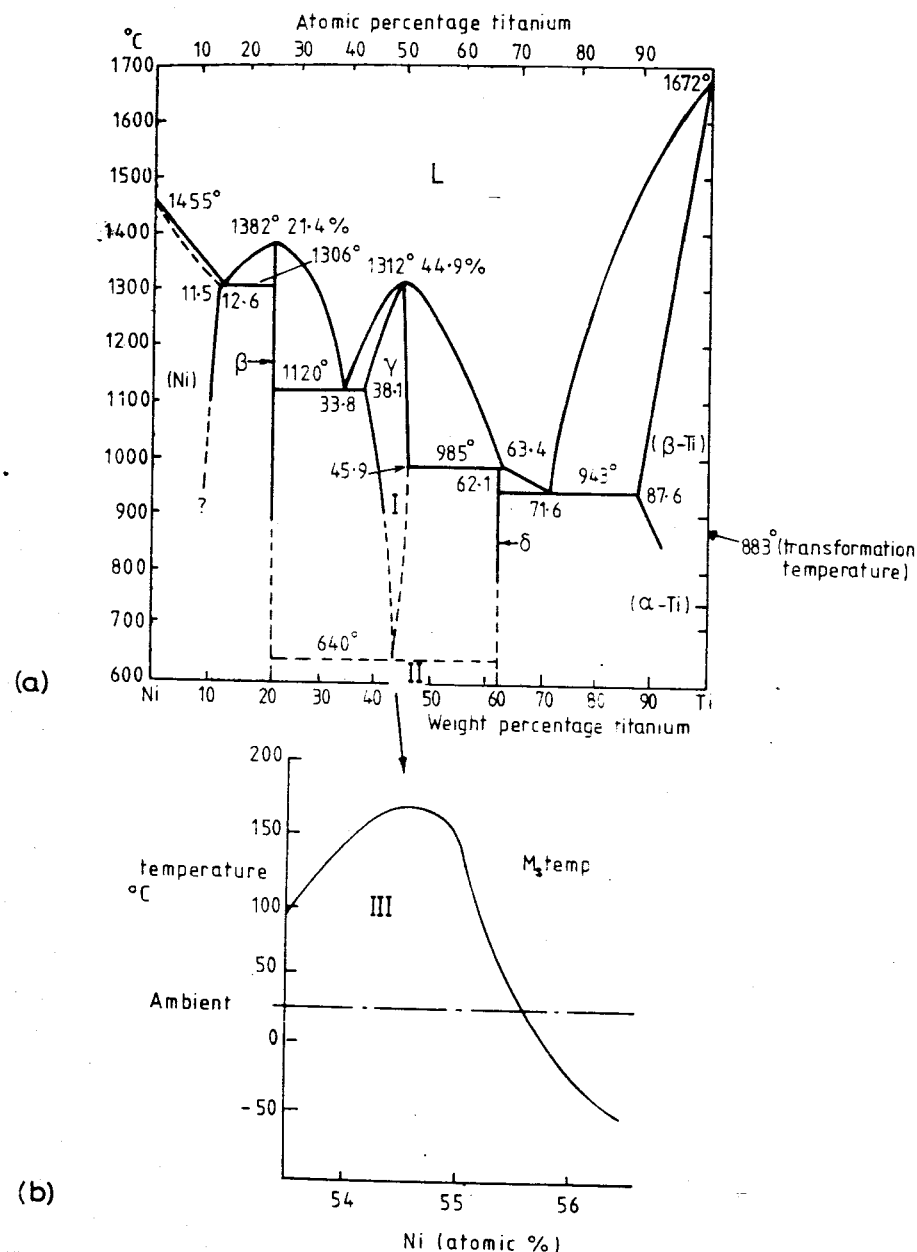


Fig. 6.34 (a) Ti-Ni phase diagram (after D.R. Hawkins in *Metals Handbook*, 8th edn., Vol. 8, American Society for Metals, 1973, p. 326); (b) M_s temperature as function of Ni content.

e.g. toys, self-erecting space antennae, special tools, self-locking rivets, etc.

Microstructure; very fine twinned martensite.

Comments: This transformation is interesting for two reasons: firstly it involves a diffusionless transformation from one ordered structure to another. It is of course fundamental to this type of transformation that if the austenitic phase is ordered, the martensitic product must also be ordered. Secondly, the mode of the transformation is such that very extensive deformation (up to 16%) can occur as a *thermoelastic* (non-plastic) martensitic shear mechanism, i.e. the transformation is *reversible*. Although this essentially involves an alloy of nominally stoichiometric composition, small additional increases in Ni (max ~55.6 wt%) can be tolerated. This increase in Ni content has the important effect of decreasing the M_s temperature. The M_s temperature as a function of Ni content is shown in Fig. 6.34b. However, it is advisable not to exceed 55.6 wt% Ni, to avoid the precipitation of the $TiNi_3$ phase. To avoid this problem, small amounts of Co can be added virtually on a 1 : 1 basis as a substitute for Ni.

While the martensitic phase is described as a complex CsCl ordered lattice, as a first approximation the transformation is related to the type: bcc $\rightleftharpoons$ hcp. The habit plane of the twinned martensite plate is irrational and close to $[551]_{bcc}$ (Sinclair and Mohammed, 1978)²⁰.

Another interesting feature of this transformation is that it appears to bring about an abrupt change in Young's modulus and yield strength. This change in Young's modulus also results in distinct changes in the modulus of resilience, the 'damping' properties of the material being much greater in the martensitic (TiNi III) form.

The large amount of deformation that this alloy can undergo due its special transformation characteristics is utilized commercially. For example, it may be "formed" in a fully reversible way simply by deforming *below* the M_s temperature. Subsequent heating above the transition temperature then changes the deformed (sheared) structure back to its original form. This unique feature of alloys such as this has made them known as: "*memory metals*". There are a number of so-called memory metals known today³⁷, although none so commercially useful as TiNi alloys.

References

1. J.W. Christian, *Theory of Transformations in Metals and Alloys*, Pergamon, Oxford, 1965.
2. Z. Nishiyama, *Martensitic Transformation*, Academic Press, 1978.
3. A.H. Cottrell, *Theoretical Structural Metallurgy*, Edward Arnold, 1962.
4. For a comprehensive review see: C.M. Wayman, *Introduction to the Crystallography of Martensitic Transformations*, Macmillan, 1964.
5. E.C. Bain, *Trans AIME*, **70**:25 (1924).

6. Z. Nishiyama, *Martensitic Transformation*. Academic Press, 1978, p. 232.
7. R. Bunshah and R.F. Mehl, *Trans. AIME*, **197**:1251 (1953).
8. R.E. Cech and D. Turnbull, *Trans. AIME*, **206**:124 (1956).
9. R. Huizing and J.A. Klostermann, *Acta Metallurgica*, **14**:1963 (1966).
10. K.E. Easterling and P.R. Swann, *Acta Metallurgica*, **19**:117 (1971).
11. C. Zener, *Elasticity and Anelasticity of Metals*, Univ. of Chicago Press, 1948.
12. J.W. Brooks, M.H. Loretto and R.E. Smallman, *Acta Metallurgica*, **27**:1839 (1979).
13. J.A. Venables, *Phil. Mag.*, **7**:35 (1962).
14. B. Lehtinen and K.E. Easterling, *High Voltage Electron Microscopy*, P.R. Swann *et al.*, (Eds) Academic Press, p. 211 1974.
15. A.J. Bogers and W.G. Burgers, *Acta Metallurgica*, **12**:255 (1964).
16. G.B. Olson and M. Cohen, *Met. Trans.*, **7A**:1897, 1905 and 1915 (1976).
17. See, e.g., C.R. Houska, B.L. Averbach and M. Cohen, *Acta Metallurgica*, **8**:81 (1960).
18. K.E. Easterling and A.R. Thölén, *Acta Metallurgica*, **24**:333 (1976).
19. See, e.g., J.P. Hirth and J. Lothe, *Theory of Dislocations* McGraw-Hill, 1968.
20. R. Sinclair and H.A. Mohamed, *Acta Metallurgica*, **26**:623 (1978).
21. G. Thomas and B.V.N. Rao, *Proc. Conf. on Martensitic Transformations*, Academy of Sciences, Kiev, 1978, pp. 57–64.
22. J.W. Eshelby, *Proc. Roy. Soc. A*, **241**:376 (1957).
23. A. Kelly, *Strong Solids*, Clarendon, Oxford, 1966.
24. See, e.g., J.Y. Koo and G. Thomas, *Mat. Sc. and Eng.*, **24**:187 (1976).
25. A.R. Thölén, *Phys. Stat. Sol.* **60**:153 (1980).
26. P.G. Shewmon, *Transformations in Metals*, p. 337 McGraw-Hill, 1969.
27. F.C. Frank, *Acta Metallurgica*, **1**:15 (1953).
28. J. Friedel, *Dislocations*, Pergamon, 1964.
29. K.E. Easterling and A.R. Thölén, *Acta Metallurgica* **28**:1229 (1980).
30. J.R. Patel and M. Cohen, *Acta Metallurgica*, **1**:531 (1953).
31. C.M. Wayman, *Phase Transformations* Inst. of Metallurgists, London, Series 3, No 11, vol 1, pp. IV–1 to 15, 1979.
32. A.M. Hunt and D.W. Pashley, *J. Phys. Rad.*, **23**:846 (1962).
33. G.R. Speich, *Metals Handbook*, vol. 8, p. 202, ASM, 1973.
34. R.W.K. Honeycombe, *Structure and Strength of Alloy Steels*, Climax Molybdenum, 1973.
35. T. Gladman, D. Dulieu, I.D. McIvor, *Microalloying 75*, Session 1, p. 25 Metals Soc., London, 1975.
36. Å. Holmström, P. Öström and R. Lagneborg, *Inst. Metals Res. Rep.* No. IM 1423 (1980).
37. See e.g. J. Perkins (Ed.), *Shape Memory Effects in Alloys*, Plenum Press, 1978.

38. See, for example, M. Hillert, 'Calculation of Phase Equilibria', Chapter 5 in *Phase Transformations*, American Society for Metals, Ohio, 1970.

Further Reading

- J.W. Christian, *The Theory of Transformations in Metals and Alloys*, Pergamon, 1965.
- R.W.K. Honeycombe, *Steels—Microstructure and Properties*, Edward Arnold, London, 1981.
- L. Kaufman and M. Cohen, 'Thermodynamics and kinetics of martensitic transformations', *Progress in Metal Physics*, Vol. 7, pp. 165–246, Ed B. Chalmers and R. King (Eds.) Pergamon, 1958.
- Z. Nishiyama, *Martensitic Transformation*, Academic Press, 1978.
- F.B. Pickering, *Physical Metallurgy and the Design of Steels*, Applied Science Publishers, 1978.
- C.M. Wayman, 'Phase transformations nondiffusive', Chapter 15 in *Physical Metallurgy*, 2nd edition, R.W. Cahn and P. Haasen (Eds.), North-Holland, 1983.
- C.M. Wayman, *Introduction to the Crystallography of Martensitic Transformations*, Macmillan, New York, 1964.
- A.K. Sinha, *Ferrous Physical Metallurgy*, Butterworths, London, 1989.

Exercises

- 6.1 Use free energy–composition diagrams to illustrate the driving force for the Fe–Ni martensitic transformation at $T > T_0$; $T = T_0$; $T = M_s$. Show how this chemical free energy can be estimated if the undercooling is known. Explain why the driving force for the nucleation of martensite at the M_s temperature is independent of carbon concentration in Fe–C steels.
- 6.2 What are the possible non-chemical energy terms in the martensitic transformation? Derive equations for the critical size and volumes of a martensite nucleus using classical nucleation theory. What evidence is there that martensite nucleates heterogeneously?
- 6.3 Evaluate Equations 6.8, 6.9 and 6.10 for Fe–C martensite assuming $\Delta G_v = 174 \text{ MJ m}^{-3}$, $\gamma = 20 \text{ mJ m}^{-2}$, $s = 0.2$, $\mu = 80 \text{ GN m}^{-2}$.
- 6.4 Give an exact definition of the habit plane of martensite. Describe how this habit plane might be measured experimentally. Give possible reasons why there is so much scatter of habit plane measurements in a given sample.
- 6.5 In the phenomenological approach to martensitic transformations there

are two *different* but *equivalent* ways of producing the lattice invariant shear. Show exactly what is meant by this. What is the experimental proof of both types of shear?

- 6.6 Draw a diagram to illustrate Bain's homogeneous deformation model for the fcc $\rightarrow$ bcc diffusionless transformation. Assuming $a_\gamma = 3.56 \text{ \AA}$ and $a_\alpha = 2.86 \text{ \AA}$, and that c/a for martensite is 1.15 calculate the maximum movement experienced by atoms during the transformation. Assume that $c/a = 1.1$.
- 6.7 What are the essential differences in martensite nucleation models based (a) on changes at the core of a dislocation; (b) on dislocation strain field interaction? Discuss the advantages and disadvantages of both models in terms of the known characteristics of martensitic transformations.
- 6.8 Give possible reasons why the habit plane of martensite changes as a function of alloying content in steels and Fe-Ni alloys. What factors influence the retention of austenite in these alloys?
- 6.9 What is the role of austenitic grain size in martensitic transformations? Is austenitic grain size important to the strength of martensite? What other factors are important to strength and toughness in technological hardened steels?
- 6.10 Suggest possible alloying and heat treatment procedures needed to design the following steels: (a) a quenched and tempered steel; (b) a dual phase steel; (c) a maraging steel; (d) a TRIP steel.
- 6.11 How would you characterize the unique properties of alloys which can be utilized as 'memory metals'. How would you design a TiNi alloy for use as, e.g., a self-locking rivet? Give instructions on how it is to be used.

Solutions to Exercises

Compiled by John C. Ion

Chapter 1

$$1.1 \quad C_p = 22.64 + 6.28 \times 10^{-3} T \text{ J mol}^{-1} \text{ K}^{-1}$$

$$\text{Entropy increase, } \Delta S = \int_{T_1}^{T_2} \frac{C_p}{T} dT$$

$$\begin{aligned} \Delta S_{300-1358} &= \int_{300}^{1358} \frac{22.64 + 6.28 \times 10^{-3} T}{T} dT \\ &= \frac{1358}{300} [22.64 \ln T + 6.28 \times 10^{-3} T] \\ &= \underline{40.83 \text{ J mol}^{-1} \text{ K}^{-1}} \end{aligned}$$

1.2

