

## *Foreword*

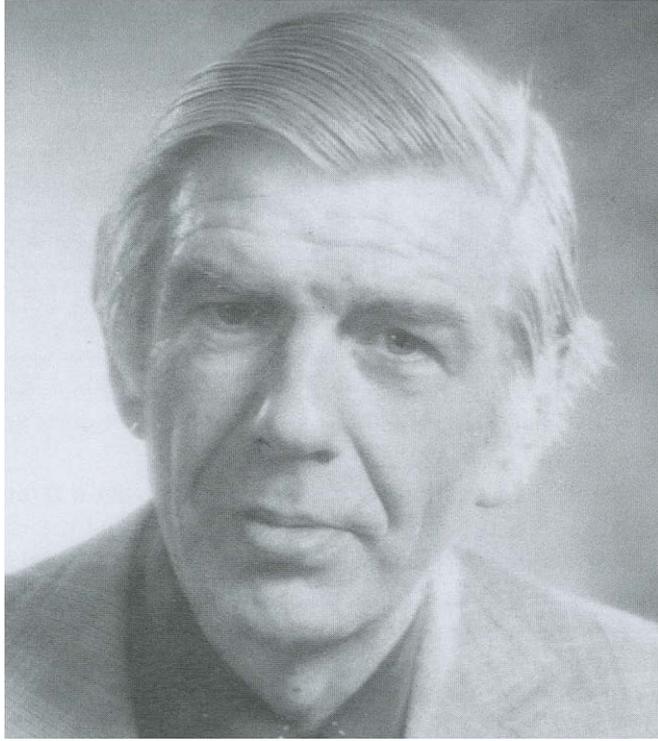
Jack was so quick to give credit to others for their work that he would not have wanted this edition of his book to be printed without thanks to his colleagues. Since he died while the book was in the final stages of production, we are unable to pass on his gratitude to the many people whom he would have wished to acknowledge.

On his behalf, I should particularly like to thank Professor Brian Cantor and Professor Adrian Sutton for taking on the considerable task of expediting the publication of this updated edition of Jack's lifetime's work. I should also like to thank everyone else who helped – having grown up with Jack's handwriting from the first letter he wrote to me while I was still at school, I appreciate their efforts.

Jack died unexpectedly in the Department he helped to create doing the work he loved. He only left two instructions, given as asides on different occasions. He wanted no Memorial Service and he wanted the last two stanzas from the Swinburne poem that he quoted in his first preface to be read at his funeral.

In the first edition he quoted the lines: "even the weariest river winds somewhere safe to sea" while complimenting his mentor, Professor Hume-Rothery on his expertise at writing books. In this one, let these lines stand as a tribute to Jack's dedication to this work throughout the course of his long and distinguished career.

Maureen Christian  
Autumn 2002



Professor John Wyrill 'Jack' Christian  
1926-2001

## *Technical Notes*

Professor Christian's classic textbook 'The Theory of Transformations in Metals and Alloys' was first published in 1965. The first part of the book was revised and re-published in 1975 and again in 1981. The present two-part set represents the long-awaited full revision of Parts I and II.

Professor Christian died on 27 February 2001 before finalising the proof reading, which included a number of technical matters such as updating the references to figures and tables, the cross references between Parts I and II, and the subject, author and symbol indexes.

Pippa Gordon helped Professor Christian with the preparation of the manuscript.

John Ball, Harry Bhadeshia, Brian Cantor, Alfred Cerezo, John Hunt, John Martin and Adrian Sutton helped to complete the proof reading.

Lorna Canderton, Henri van Dorssen, Caroline Kraaijveld, Julia Wallace and Vicki Wetherell helped to complete the publishing process.

Gratitude is expressed to the freelancers involved in this project.

Separate subject indexes appear in both Parts I and II. The author index and index of symbols from the 1981 edition appear in Part II only. Prefaces to earlier editions have been included for completeness. The problem set mentioned in the 1975 Preface to the Second Edition has not been included.

Autumn 2002

## *Preface to the Second Edition (of Part I)*

The arrangement of chapters and sections is unchanged but the book is now divided into two parts, the first of which covers the general theory whilst the second is concerned with descriptions of specific types of transformation. My original intention was to make only limited revisions and additions but I eventually decided that very extensive rewriting was required in view of the large number of papers published in the last ten years. Part I of the new edition is consequently about 50% longer than the corresponding chapters of the first edition, and is appearing after a lengthy delay, for which I apologize. However, I hope that all the major developments of the interim period have been incorporated, so that the edition may perhaps serve both as a text book and as a reference source up to the level of active research.

An increased symbol list has necessitated some minor changes of notation (especially  $Q$  instead of  $Z$  for partition functions) and numerical quantities are now given in SI units. In response to several requests, I have included a few selected problems at the end of Part I; these are of a very variable standard, but I hope that teachers and students may find some of them useful\*.

I should like to thank many friends and colleagues for their kind comments and for notes of mistakes or misprints in the first edition. New errors will inevitably have appeared in this volume, and I shall be very grateful if they are also brought to my attention.

*Oxford*  
*December 1974*

J.W. CHRISTIAN

\*Please see Technical Notes on page vii.

*Preface to the Second Edition, Reprinted with  
Corrections (of Part I)<sup>†</sup>*

This reprint with corrections appears, to my considerable embarrassment, before Part II of the second edition, the preparation of which is taking very much longer than originally hoped. A number of minor misprints and errors have been corrected and I have also added a few brief references to work published in the last five years. The text on page 224–5<sup>‡</sup> follows very closely the description given in an early and authoritative review article by Dr. L. Guttman, but owing to an oversight on my part, his work was not previously acknowledged. I am very happy to have an opportunity to correct this omission.

*Oxford*  
1981

J.W. CHRISTIAN

<sup>†</sup>Originally called 'Addendum' in 1981 edition.

<sup>‡</sup>Now on pages 226–7.

## *Preface to the First Edition*

Suitable thermal or mechanical treatments will produce extensive rearrangements of the atoms in metals and alloys, and corresponding marked variations in physical or chemical properties. In this book, I have tried to describe how such changes in the atomic configuration are effected, and to discuss the associated kinetic and crystallographic features. The most drastic rearrangements accompany phase transformations, and the atomic processes leading to changes in the phase structure of an assembly form the main topic of the book. There are, however, many changes in the solid state which are not phase reactions, but which depend on the same atomic processes. I have also classed these changes as “transformations” since they are distinguished from phase transformations only by the nature of their driving forces. In terms of mechanism, kinetics and crystallography, it should be possible to give a consistent description of all these carried phenomena.

The subject matter so defined includes virtually all major kinetic effects in physical metallurgy and solid state physics, with the notable exception of plastic deformation by slip. A good case could be made for including this also in a unified treatment with the other phenomena, but it is so vast a subject that it would have doubled the size of an already large book, and would have been beyond my competence to attempt. When we think about transformations in the solid state, however, the theory of plastic deformation is never far in the background, and the properties of lattice defects are equally vital in both cases. Moreover, some types of phase transformation may also be considered as modes of deformation, so that mechanical twinning, for example, is included in this book because of its close relation to martensitic transformation.

The book does not deal with the static problems of metal physics, notably the prediction of the equilibrium configuration from the known properties of the atoms. Nevertheless, I felt that a self-consistent description of the kinetic theories is not possible without a reasonably detailed description of the (usually crude) models used to describe the equilibrium or metastable state. The book thus contains what I hope are rather complete discussions of such topics as lattice geometry, point defects, dislocations, stacking faults, grain and interphase boundaries, solid solutions, diffusion, etc., and there are short accounts of more controversial topics such as the thermodynamics of the steady state. These introductory chapters form an appreciable fraction of the whole book, and unlike the kinetic sections, they cannot be justified on the grounds that no comparable

description exists in other texts. I have two excuses for including them here. One is that the existing accounts, excellent though they are, frequently lack either the emphasis or the detail required for the present discussion of transformation problems. Thus I have found no adequate descriptions (except in original papers and specialized review articles) of such vital topics as the geometry of finite homogeneous deformation, the structure of a crystal surface, the properties of twinning dislocations, or the elastic energy of a constrained transformation. My second excuse is best illustrated by a quotation from Dr. Ziman's preface to his own recent book:<sup>†</sup> "There is need for treatises covering, in reasonable detail, up to the level of active research, the major branches into which the subject has divided." After attempting such a task, I am very conscious of my own inadequacies, but even with imperfect optics, there is something to be said for the wide-angle lens. In a very tentative sense, I hope that the present volume may find some use as an advanced account not only of phase transformation theory, but of a much wider field of physical metallurgy.

Throughout the book, I have tried to emphasize the theoretical description of the various phenomena, and experimental results are selected more for their ability to illustrate a particular argument than because they are the best data available. As in other branches of solid-state research which are sensitive to defect configurations and interactions, much of the recent active research work is intended to give a detailed description of the observations made on a particular metal or alloy, and this presents some difficulties to the writer of a general text. I have tried to solve these by giving at least one detailed description of each important effect. Thus the theoretical coverage is intended to be reasonably complete, and I may be justly blamed if it is not; the experimental results are only examples from a much wider range. The literature has been thoroughly surveyed up to the end of 1961, and work from some of the more important papers up to September 1962 has also been included.<sup>‡</sup> However, my main aim has been to present an account of the established body of theory, so except in one or two special cases (e.g. Cahn's theory of spinodal decomposition), it has not seemed particularly important to include all the latest papers. After a period of very rapid development, it now seems likely that we are entering a stage of consolidation in this subject, so the present time seems well suited to the publication of a more detailed account than is to be found in any of the collections of review articles.

In writing this book, my ideal reader has been the graduate student just beginning research, although I hope that much of it may be useful at an undergraduate level. Some knowledge of elementary crystallography, chemical thermodynamics, and statistical mechanics is assumed, but relevant parts of these subjects are developed almost *ab initio*. The mathematical development should not give much trouble to an undergraduate reading for an honours degree in physics, but the metallurgical student may be a little alarmed by the extensive use of vector analysis and matrix algebra. The matrix method is so powerful in handling the crystallographic problems, and especially the compatibility between changes in the unit cell, distribution of defects, and changes in macroscopic shape, that it would have been inept not to have used it. However, I have given parallel geometrical

<sup>†</sup>*Electrons and Phonons*, Clarendon Press Oxford, 1960.

<sup>‡</sup>A few references to important papers published up to mid-1964 are being added in proof.

illustrations of most of the matrix manipulations, so that the physical ideas may be followed even if the details of the calculation appear obscure. Only elementary matrix algebra is used, and the student to whom this is new could learn sufficient about it from any good textbook in a few hours. The notation is explained rather fully in the text, as are some standard operations such as the procedure for finding the eigenvalues of a matrix.

The only books which have previously covered any appreciable part of the subject matter are Volmer's *Kinetik der Phasenbildung* and Frenkel's *Kinetic Theory of Liquids*. I have greatly enjoyed reading these two classical texts, but they were written before the importance of defects in transformations was fully appreciated, and the main sources for this book are therefore individual papers and review articles. These are listed separately in the references, but I mention here my indebtedness especially to the various volumes of *Progress in Metal Physics* (now *Progress in Materials Science*), edited by B. Chalmers and R. King, and of *Solid State Physics*, edited by F. Seitz and D. Turnbull, and also to *Phase Transformations in Solids* (Wiley, 1951) and *The Mechanism of Phase Transformations in Metals* (The Institute of Metals, 1956). Many authors and publishers have kindly given permission for the reproduction of figures; the sources are individually acknowledged in the text. I am especially grateful to those friends who kindly supplied prints of optical or electron micrographs. Chapter 20 is based largely on unpublished material from Dr. A.G. Crocker's Ph.D thesis, and I am very grateful to him and to Professor Bilby for allowing me to use it.

I have not shown the text to any of my colleagues, so that I have sole responsibility for any misrepresentations or errors of fact or judgement which may be detected, and I shall be very grateful if these are brought to my notice. On the other hand, I am extremely conscious of the help I have received from very many friends in numerous discussions of specific points, and I should like to mention especially B.A. Bilby, R. Bullough, J.W. Cahn, M. Cohen, A.G. Crocker, A. Hellawell, R.F. Hehemann, M. Hillert, D. Hull, A. Kelly, P.M. Kelly, D.S. Lieberman, J.K. Mackenzie, J. Nutting, W.S. Owen, T.A. Read, C.S. Smith, G.R. Speich, J.A. Venables, C.M. Wayman and M.S. Wechsler. I owe an especial debt to Dr. Z.S. Basinski, whose ideas contributed so much to my first understanding of the problems of martensite crystallography. The book has been written over many years at Oxford, and throughout this time I have had the benefit of Professor Hume-Rothery's help and encouragement. He is so good at writing books himself that he can probably not understand why it has taken me so long, but he will be relieved to discover that even the weariest river winds somewhere safe to sea. The final preparation of the manuscript has been undertaken during a sabbatical leave spent at Case Institute of Technology, and I should like to express my warm gratitude to Professor A.R. Troiano and his staff for their hospitality and many kindnesses.

Cleveland, Ohio  
Oxford, England

J.W. CHRISTIAN

## CHAPTER 1

### *General Introduction*

#### 1. CLASSIFICATION OF TRANSFORMATIONS

An assembly of atoms or molecules which has attained equilibrium under specified external constraints consists of one or more homogeneous and physically distinct regions. The regions of each type may be distinguished by a common set of parameters defining such intrinsic properties as density, composition, etc., and they constitute a phase of the assembly. Two phases are distinguishable if they represent different states of aggregation, different structural arrangements in the solid, or have different compositions. This book deals mainly with changes in the phase structure of metals and alloys when the external constraints are varied.

In a complete discussion of the theory of phase changes, the following two questions must be considered:

- (1) Why does a particular phase change occur?
- (2) What is the mechanism of the transformation?

To answer the first question we must investigate the properties of different arrangements of a given assembly of atoms or molecules and thus attempt to find the equilibrium configuration. The formal theory of equilibrium, developed by Willard Gibbs, is expressed in terms of macroscopic thermodynamic parameters, and the problem reduces to the evaluation of these quantities from the properties of isolated atoms or molecules. Ideally, this can be tackled from first principles by quantum mechanics, but since the wave equation to be solved contains  $\sim 10^{22}$  variables, it is clear that drastic simplifications are required. Much work has been devoted to finding suitable methods of calculation but the results were not encouraging until the early seventies, and in particular it was not generally possible to predict the relative stabilities of different crystal structures. This is not surprising when it is remembered that heats of transition in metals which exist in more than one solid equilibrium form are only of the order of one percent of the binding energy of the solid as measured by the latent heat of vaporization. The calculation of the binding energy was itself a very difficult problem, and in various approximate methods, it was obtained as the sum of several terms, the uncertainty in any of which was usually greater than the net difference in energy between alternative structures. It is true that in some favourable cases, it was possible to provide

plausible explanations why one structure should be more stable than another, but these explanations were essentially semi-qualitative, being based on physical models rather than detailed calculations.

In the last twenty years, all this has changed. The advent of what is termed “density functional theory” has enabled calculations of quantum states to be made from first principles so that in the words of Sutton and Balluffi (1995) “it is now possible to calculate using the atomic number and some local density functional as the only input, the following properties more or less routinely: the stability of elements and alloys in various observed or hypothetical crystal structures (as a function of atomic volume or of some affine transformation such as a shear), equilibrium lattice constants, stacking fault and surface energies, elastic constants, phonon dispersion relations, band structures and Fermi surfaces (based on Kohn-Sham eigenvalues), and the relaxed atomic and electronic structures and energies of periodic interfaces with small planar unit cells”. Hohenberg and Kohn (1964) proved that all aspects of the electronic structure of a system of interacting electrons in an external potential and a non-degenerate ground state are completely determined by the electron charge density. This theorem enables the major difficulty of the number of variables to be reduced to three. The ground state energy is a unique functional of the charge density, and the only difficulty is that this functional is unknown since it corresponds to the solution of the full many electron problem. However, several useful approximations to the functional have been proposed and have led to the remarkable success in predictions made above. Later Kohn and Sham (1965) derived a set of self-consistent one particle equations for the description of electronic ground states.

A major disadvantage of the density functional method is that it invariably requires substantial computational resources. For this reason, many calculations are made with other more approximate methods. Before the development of density functional theory, most atomistic calculations were based on the assumption of pair potentials. This is a poor approximation for transition metals and may be unreliable even for metals with primarily s-p bonding. There is a hierarchy of simpler methods which do not use full density functional theory. Tight binding approaches treat the electrons explicitly but in a simplified manner that can be justified by well defined approximations to density functional theory (Sutton et al, 1988). The tight binding approach describes directional bonding through the angular character of atomic orbitals. More empirical models for metals attempt to fold the influence of the electrons into effective interatomic potentials that are multi-atom rather than pair-wise in nature. Among these, the “embedded atom” potential of Daw and Baskes (1984) and the “Finnis Sinclair” potentials (Finnis and Sinclair, 1984) have been widely used. In these models metallic cohesion is described by an interatomic potential that is dependent on the local coordination number, but which does not capture any directional bonding in the system. Empirical potentials for the bond order in covalently bonded insulators such as silicon and diamond have been developed ( Tersoff , 1988) and widely used. The bond order is related to the strength of the bond. Pettifor (1989) has derived an expansion for the bond order in real space, which is exact (Pettifor and Aoki, 1991).

The second fundamental question asks how a transformation occurs. The methods of classical thermodynamics are now of more limited application, since phase transformations

are "natural changes". The appropriate theories are essentially kinetic theories, and some model of the atomic processes involved is implicit in any treatment. The models to be used are of two kinds. We first require an acceptable description of the "ideal" structure, using approximations for the interatomic forces to make them amenable to mathematical treatment. This is the problem already mentioned, but fortunately the kinetic properties are not so sensitive to the nature of the binding forces, so that assumptions which would be unwarranted in a study of equilibrium properties may be used with rather more confidence. We have next to consider the deviations from the ideal structure which occur in real solid materials. Vacant lattice sites, interstitial atoms, dislocations, stacking faults, and grain boundaries are comparatively unimportant in a description of the equilibrium state, but their presence may be fundamental to the process of transformation. Properties which depend sensitively on the detailed structure of the assembly are sometimes called structure sensitive.

In this book we shall be concerned mainly with the mechanism of phase transformations, and the reasons for transformation will not be examined in any thorough manner. In order to deal efficiently with the second question, however, the models mentioned above must be developed in some detail, so that we shall also consider some aspects of equilibrium theory. This is necessary, partly to make our treatment reasonably complete and self-contained, and partly because many of the topics which are important in discussing transformations are left out of the usual descriptions of equilibrium properties. It should be obvious from the above discussion that the models to be described refer almost exclusively to the solid state. For the vapour state, the kinetic theory of gases provides an adequate approximation, and there is no readily visualized model which gives a satisfactory description of the liquid state. The first part of this book is thus devoted mainly to an account of the solid state in metals and alloys, and the second part to the ways in which transformations involving any of the states of matter may occur.

Any phase transformation requires a rearrangement of the atomic structure of the assembly. In the solid state, similar rearrangements take place during processes which are not phase reactions, for example during the recrystallization of a deformed metal or its subsequent grain growth. Such reactions are distinguished by their driving forces;<sup>†</sup> the atoms take up new relative positions under the influence of strain energy, surface energy, or external stress, and not because the free energy of one arrangement is inherently lower than that of the other. The atomic mechanisms involved in all these changes, however, are closely related, and it is advantageous to treat them together. The word "transformation" will thus be used in a general sense to mean any extensive rearrangement of the atomic structure. The definition is intended to exclude mechanical deformation by slip, which (in principle) only alters the atomic arrangement by translation of part of the structure over the remainder. Deformation twinning is included because of the highly ordered nature of the rearrangement and its close relation to one type of phase transformation.

<sup>†</sup> In any natural process the free energy of a closed assembly in thermal contact with its surroundings decreases. The difference in initial and final free energies provides a driving energy or thermodynamic potential for the change; this is often loosely described as a "driving force".

The driving force for any transformation is the difference in free energy (usually Gibbs free energy) of the initial and final states, and is thus determined by thermodynamic parameters appropriate to large regions of the phases concerned. The mode of transformation is very dependent on the effect of small fluctuations from the initial condition, and in particular on the question of whether such a fluctuation raises or lowers the free energy. A metastable assembly is resistant to all possible fluctuations, and any transformation path must pass through states of higher free energy. Conversely, if any infinitesimal fluctuation is able to lower the free energy, the initial condition is unstable, and there is no energetic barrier to transformation along a path represented by this fluctuation. A truly unstable phase only has a transitory existence, but it may also be possible to obtain a phase in such a condition that the only barrier to transformation is that limiting atomic movement or diffusion. Such an assembly may also be described as unstable, although it decomposes at a finite rate determined by the diffusion rate.

In considering the problem of stability, Gibbs distinguished two different kinds of fluctuation, namely those corresponding to fairly drastic atomic rearrangements within very small localized volumes, and those corresponding to very small rearrangements spread over large volumes. An example of the first kind of fluctuation is the formation of a very small droplet of liquid in a supersaturated vapour, whilst an example of the second kind is the development of a periodic variation of composition with a very long wavelength within an initially homogeneous solid solution. Much confusion in discussions of the initiation of phase transformations in supersaturated solid solutions may be attributed to a failure to distinguish clearly between these two kinds of fluctuation.

Most of the transformations with which we shall be concerned are heterogeneous; by this we mean that at an intermediate stage the assembly can be divided into microscopically distinct regions of which some have transformed and others have not. Transformation thus begins from identifiable centres in the original phase, a process which is called nucleation. The classical theory of the formation of a nucleus is formulated in terms of the first type of fluctuation, and it may be shown that any assembly is stable to such fluctuations on a sufficiently small scale. The reason for the free energy barrier is usually expressed in the form that the negative free energy change resulting from the formation of a given volume of a more stable phase is opposed by a positive free energy change due to the creation of an interface between the initial phase and the new phase. As the volume of the region transformed decreases, the positive surface term must eventually dominate the negative volume term, so that the whole free energy change becomes positive. Clearly these macroscopic concepts are not really applicable to very small fluctuations, and any separation of the net free energy change into volume and surface contributions is arbitrary. However, the formalism is useful in deriving quantitative expressions for the nucleation rate, provided due caution is exercised in identifying parameters introduced into the nucleation theory with those derived from measurements on bulk phases, and it also illustrates the conditions under which the energy barrier to nucleation might disappear.

For localized fluctuations of any one kind, there is a critical size of nucleus at which the free energy barrier is a maximum, and the magnitude of this maximum increase in free energy determines the rate of nucleation according to the classical theory. If the surface

energy decreases, the critical nucleus size and the height of the energy barrier both decrease, and more and more nuclei will be formed in a given volume in a given time. In the limit, the nucleation barrier disappears altogether when the surface energy becomes zero, or at least very small, and the original phase becomes unstable. The transformation will then be homogeneous, taking place in all parts of the assembly simultaneously. The condition for a homogeneous transformation thus appears to be a zero, or near zero, surface energy, and this implies that there must be no abrupt change at the boundary between two phases. In some order-disorder transformations, it is possible that this condition is readily approached, so that homogeneous or near homogeneous transformation is possible. The surface between two phases of similar structure but different composition and lattice parameter may also have a low effective energy if the transition from one phase to the other is spread over a wide region, i.e. if there is a diffuse rather than a sharp interface. However, a diffuse interface is clearly incompatible with a localized fluctuation, so that in attempting to remove the energy barrier from the first kind of fluctuation we have essentially converted it into the second kind of fluctuation. Homogeneous transformations without nucleation are thus generally possible only in certain assemblies which are unstable with respect to the second kind of fluctuation. The most important examples of such homogeneous transformations in non-fluid assemblies are probably those already mentioned, namely the decomposition of supersaturated solutions within certain limits of temperature and composition, and some order-disorder changes. It is generally impossible to avoid nucleation in transitions between solid phases with different structures, since the transition from one arrangement to another cannot then be made continuous through a diffuse boundary.

Heterogeneous transformations are usually divided into two groups, originally distinguished from each other by a different dependence of reaction velocity and amount of transformation on temperature and time. This experimental classification almost certainly corresponds to a real difference in the physical mechanisms of transformation, but experience has shown that the kinetic features cannot always be interpreted unambiguously, and many reactions of intermediate character are now recognized. The two groups are usually known as "nucleation and growth" reactions and "martensitic" reactions respectively. This nomenclature is rather unfortunate, since the formation of stable small regions of product and the subsequent growth of these nuclei may have to be treated as separate stages in both classes of transformation. There are also now known to be several examples of transformations with characteristics intermediate between those of the two categories.

In typical transformations in the first group, the new phase grows at the expense of the old by the relatively slow migration of the interphase boundary, and growth results from atom by atom transfers across this boundary. The atoms move independently and at a rate which varies markedly with the temperature. At a given temperature, the reaction proceeds isothermally, and the amount of new phase formed increases with time. Although the volume of a transformed region differs in general from its original volume, its shape is substantially unaltered except in certain plate shaped products. Such changes are largely governed by thermal agitation. The unit processes (e.g. the motion of an individual atom within the parent phase or across the interphase boundary) are similar to chemical reactions, and to a first approximation they may be treated formally by the methods of the statistical reaction rate theory.

Nucleation and growth transformations are possible in all metastable phases, and the initial or final condition may be solid, liquid, or gaseous. The second kind of heterogeneous transformation, however, is only possible in the solid state, and utilizes the co-operative movements of many atoms instead of the independent movements of individual atoms. Most atoms have the same nearest neighbours (differently arranged) in the two phases, and the net movements are such that in small enough regions a set of unit cells of the original phase is effectively homogeneously deformed into a corresponding set of cells of the new phase. The transforming regions then change their shapes, and may be recognized, for example, by the disturbances produced on an originally flat polished surface. Discrete regions of the solid usually transform suddenly with a very high velocity which is almost independent of temperature. In most cases, the amount of transformation is characteristic of the temperature, and does not increase with the time. Reactions of this kind are often called diffusionless or shear transformations, but in recent years it has become common practice to refer to them as martensitic transformations. The name is an extension of the nomenclature originally reserved for the hardening process in quenched steels.

In recent years it has become increasingly evident that many transformations do not fit easily into the above classification. Difficulties arise especially when the growth is thermally activated but nevertheless results in the type of shape change and crystallography normally associated with a martensitic transformation. Provided the shape change has been properly measured (Clark and Wayman, 1970), it implies the existence of a correspondence of lattice sites of parent and product structures (see Chapter 2), even though in some cases there has been long range diffusion, resulting in a composition difference between product phase and the phase from which it is growing. Such transformations may be called diffusional displacive (Christian, 1997). Lieberman (1969, 1970) has proposed that the various kinds of transformation which give rise to a martensitic shape change should be subdivided into ortho-, para-, quasi-, and pseudo-martensites, but these names have not been generally accepted. We feel that it is preferable to restrict the word martensitic to changes which are in principle diffusionless, and to use some more general term for all transformations, including martensite formation, which involve a lattice correspondence or partial correspondence.

The descriptive terms "displacive" and "reconstructive" transformations have a long-established usage, especially in non-metallurgical fields (Buerger, 1951) and may seem more adaptable than their metallurgical equivalents. This classification, however, is rather closely linked to the concept of pairwise atomic interactions, or bonds. An alternative division of a more general kind (Frank, 1963; Christian, 1965) is based on a sustained analogy between the different mechanisms of transformation and the ways in which soldiers and civilians respectively execute some simple task, such as boarding a train. Thus the main categories of transformation are called "military" and "civilian", but rigid classification is not required since soldiers may sometimes be out of step and civilians may sometimes form paramilitary organizations!

Boundaries in the solid state may be conveniently regarded as either glissile or non-glissile. A glissile boundary can migrate readily under the action of a suitable driving stress, even at very low temperatures, and its movement does not require thermal activation. Examples of the motion of glissile boundaries are provided by the growth of a martensite plate or of

a mechanical twin, or by the stress-induced movement of a symmetrical low angle grain boundary. In all cases, the shape of the specimen changes as the boundary is displaced, so that the movement may be regarded as a form of plastic deformation. It follows that a suitable external mechanical stress should be able to produce displacement of any glissile interface.

The remaining types of boundary can move only by passing through transitory states of higher free energy, so that the motion requires the assistance of thermal fluctuations. However mobile such a boundary may be at high temperatures, it must become virtually immobile at sufficiently low temperatures. We subdivide non-glissile boundaries into those in which there is no change of composition across the interface and those dividing regions of different composition. In the first group are any transformations from a metastable single phase to an equilibrium single phase (polymorphic changes), processes such as recrystallization and grain growth which are entirely one-phase, and order-disorder reactions. In all these examples, the rate of growth is determined by atomic processes in the immediate vicinity of the interface, and we may describe such growth as "interface controlled".

Familiar examples of growth in which there is a composition difference across a moving interface are provided by precipitation from supersaturated solid solution and eutectoidal decompositions. The motion of the interface now requires long-range transport of atoms of various species towards or away from the growing regions, so that we have to consider the diffusional processes which lead to the segregation. Two extreme cases can be distinguished in principle. In one of these we have a boundary which can move very slowly, even under the influence of high driving forces. The rate of motion will then be largely independent of the diffusion rate, and we may again describe the growth as interface controlled. The other extreme case is where the boundary is highly mobile when compared with the rate of diffusion, so that it will move as rapidly as the required segregation can be accomplished. The growth rate is then determined almost entirely by the diffusion conditions, and is said to be "diffusion controlled". Since on the average an atom will have to make many hundreds or thousands of atomic jumps in the parent phase, and only one or two jumps in crossing the interphase boundary, it is rather more likely that growth will be diffusion controlled when the composition difference is appreciable. In many diffusion controlled reactions, a linear dimension of the growing product region is proportional to the square root of the time of growth, whereas in an interface controlled reaction, it is linearly proportional to the time. Thus an interface controlled boundary of a very small product region may become diffusion controlled as the region grows larger.

In an interface controlled transformation, regions in the immediate vicinity of the interface, but on opposite sides of it, must have different free energies per atom, this being the driving potential for the growth. It follows that the temperature-composition conditions at the interface cannot correspond to equilibrium between the bulk phases, and the interface velocity will be some function of the deviation from equilibrium, as measured, for example, by the supercooling or supersaturation at the interface. In contrast to this, the deviation from equilibrium at the interface is negligible when the growth is controlled by the diffusion of matter (or energy). The overall growth rate is then a function of the difference in composition (or temperature) between regions of the primary phase near the interface

and metastable regions remote from the interface. Obviously there is a range of interface mobilities over which the growth rate cannot correspond to either extreme, and this range must be encountered if there is a transition from interface control to diffusion control as a product region grows. When comparable driving forces are needed to maintain the interface mobility and to remove or supply solute (or heat), the growth velocity will be dependent on both the local deviation from equilibrium at the interface and the gradients of composition (or temperature) in the parent phase. It is usually assumed that this range is quite small, so that the motion of most non-glissile interfaces is essentially controlled either by atomic kinetics at the interface or by diffusion.

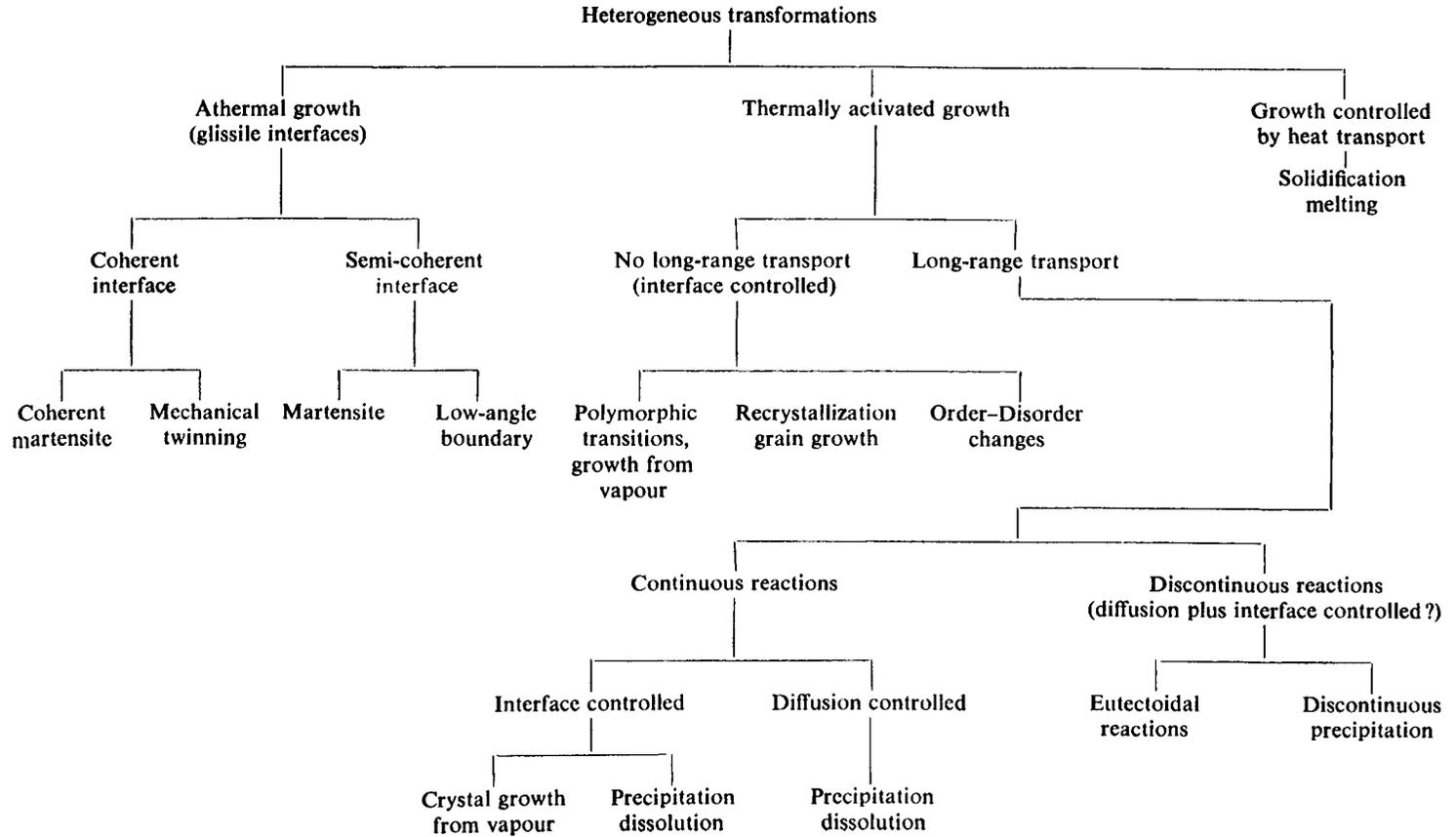
Linear growth does not necessarily imply interface control. Dendritic growth is mainly diffusion controlled, but dendrites lengthen at constant rate. Certain solid state reactions (discontinuous precipitation, eutectoidal decomposition), in which a duplex product grows into a single phase parent, also merit special consideration. The product and parent have the same mean composition, but the product consists of alternate lamellae differing in composition. These reactions have linear growth rates, but have usually been described as diffusion controlled, the operative diffusion path being either through the parent phase, or along the boundary between parent and product. In contrast to this, some theories imply that two parameters are needed to specify the growth conditions in such transformations, the interface velocity not being controlled completely by either the diffusion coefficient or the interface mobility.

In the description of growth processes, we have neglected the heat which is released or absorbed during the growth. In most transformations, the rate of growth is not limited by the rate at which this heat is supplied or removed, but this is not true when we extend the classification to reactions involving phases which are not solid. In particular, the rate of growth of a solid crystal from the melt, or from a liquid solution, is often controlled by the energy flow conditions, and this is also true of the reverse process of melting. Solidification under conditions in which the heat can be readily removed, as in a thin-walled capillary tube, is interface controlled. Growth from the vapour phase, whether or not it involves long range transport, is almost always interface controlled. Energy flow control is always equivalent to a form of diffusion control, and is covered by including variations of temperature as well as composition in the above statements.

The relations between the various forms of transformation to be discussed in this book are shown schematically in Table I, which summarizes the growth classification we have just described. Some further subdivisions are made for convenience, and the significance of these will become apparent later. It should be obvious from the above discussion that the category in which a given type of transformation is placed will be dependent to some extent on the experimental conditions under which it is observed, so that the divisions indicated are by no means rigid.

Despite its limitations, we believe the scheme shown in Table I still gives the most useful physical picture of the interrelations among the various transformation mechanisms. This scheme is approximately the same as that adopted by the committee on phase transformations for the survey *Perspectives in Materials Research* (Cohen *et al.*, 1965), but Guy (1972) has suggested a slightly different classification in which the initial division into homogeneous

TABLE I. CLASSIFICATION OF TRANSFORMATIONS ACCORDING TO GROWTH PROCESSES



General Introduction

and heterogeneous transformations is discarded by adopting a more general definition of a phase (Guggenheim, 1967). An inhomogeneous assembly is then treated as an infinity of infinitesimal phases, and a surface is regarded as a separate phase. This formalism may have certain advantages, but we believe that the Gibbs concept of a phase is clearer, and that the distinction between homogeneous and heterogeneous transformations is useful because the nucleation process may be very important in heterogeneous transformations.

Our classification focuses attention on the growth process, but either the growth rate or the nucleation rate may be effective in determining the overall kinetics of a heterogeneous transformation. The net transformation rate will depend mainly on the slower of the two stages, becoming virtually zero if either the nucleation rate or the growth rate is sufficiently slow. We should thus consider whether or not there are differences in the physical mechanisms leading to the formation of nuclei in particular transformations; if there are such differences, a classification of transformations based on the nucleation stage will be equally valid. We shall find, in fact, that there are various ways in which nuclei may be formed, but these are not readily distinguished by experimental criteria. The growth classification is convenient largely because most reactions in the solid state are greatly influenced by the growth mechanism.

In some transformations, there is no doubt that thermal fluctuations resulting in atomic rearrangements are responsible for the appearance of stable nuclei of a new phase. Such fluctuations may occur randomly throughout the volume of the assembly, or at preferred sites where impurities or structural defects act as catalysts for the phase change. Many solid assemblies contain defects which are not in thermal equilibrium with the structure but are "frozen in". Non-equilibrium defects of this kind may be able to form suitable nuclei for some changes without the aid of thermal agitation.

The classical theory of random (bimolecular) fluctuations provides an adequate description of nucleation phenomena when both phases are fluid (e.g. the condensation of a vapour). This theory has also been applied to reactions in condensed phases, and despite many difficulties and uncertainties, it probably gives a qualitatively correct description of most thermally activated nucleation processes. Since the distinction between martensitic and nucleation and growth reactions has been made on the basis of whether or not growth is thermally assisted, it is natural to enquire whether the same distinction applies to the nucleation stage. There is evidence that thermal nucleation is not needed for some martensitic transformations, but it is certain that thermal agitation plays a role in the nucleation stage of other such reactions. It does not follow from this that the classical theory of homogeneous nucleation applies, since thermal vibrations (or zero point energy) may only be helping to overcome initial obstacles to the growth of existing defects. An analogy is sometimes made with deformation processes, where thermal vibrations enable existing glide dislocations to overcome obstacles to their motion, although they are quite unable to form new dislocations in regions of perfect lattice.

These remarks apply equally to any nucleation process, including those of many nucleation and growth reactions. In suitable conditions, it is conceivable that reactions which require thermally activated migration for the growth process may begin from essentially athermal nuclei formed from existing defects. Such changes, although still classed as nucle-

ation and growth transformations for convenience, do not require nucleation in the classical sense. Conversely, when thermally assisted nucleation becomes the dominant factor in determining the overall transformation rate in a martensitic reaction, the kinetic features naturally resemble those of most nucleation and growth reactions. For this reason, the change of shape in the transformed region is probably the most definitive experimental observation in identifying a martensitic transformation.

We conclude from this discussion that when transformation does begin from particular centres in the assembly, the nucleation and growth processes are independent, in the sense that either or both may be thermally activated or athermal. We should also emphasize again that there are intermediate categories of transformation, which may require atomic diffusion, but result in a change of shape in the product regions.

## 2. CHARACTERISTICS OF NUCLEATION AND GROWTH TRANSFORMATIONS

The velocity of transformation in nucleation and growth reactions may be dependent both on the rate at which stable nuclei form and on their subsequent growth rates. In some transformations, the activation energy for nucleation is the only important rate limiting factor, but in condensed phases the activation energy for atomic migration or diffusion is usually equally important.

The general characteristics of these reactions may be summarized as follows.

(1) *Dependence on time.* At any temperature, the amount of transformation increases with time until a state of minimum free energy for the assembly is reached. In practice, however, transformation at some temperatures may be so slow that it cannot be detected in any observable period of time. The isothermal transformation laws are considered in more detail in Section 4 below.

(2) *Dependence on temperature.* If given sufficient time, the transformation will, in principle, continue until complete. The amount of transformation does not, therefore, depend on temperature, except in the trivial sense that the equilibrium state is itself a function of temperature. The velocity of transformation varies enormously, and for any transformation a temperature can be found below which the change proceeds at a negligible rate. For homogeneous changes, the reaction velocity increases approximately exponentially with the temperature over the whole range in which transformation may occur. The theory of rate processes of this kind is treated in Chapter 3. For nucleation and growth reactions, however, the rate of reaction becomes zero at the thermodynamical transformation temperature, since the energy required for nucleation is then infinite. Consider a transformation which occurs on cooling. As the temperature is lowered, the free energy of formation of a critically sized nucleus decreases much more rapidly than does the available thermal energy. The probability of nucleation may thus increase rapidly with decreasing temperature, at least until very low temperatures are reached. In condensed phases, however, the formation energy of the critical nucleus constitutes only part of the required activation energy, and there may be additional terms representing the energy needed for an atom to cross the

interphase boundary or for interdiffusion of the species, where a composition change is involved. Since the activation energy for this type of process is nearly temperature independent, the rate of growth of both sub-critical and super-critical nuclei decreases as the temperature is lowered. The decreasing formation energy and decreasing growth rate produce a characteristic dependence of reaction velocity on temperature, in which the rate increases to a maximum at a temperature below the equilibrium transformation temperature, and then decreases again.

This type of variation often makes it possible to obtain an alloy in a thermodynamically metastable state. This can be done if the specimen can be cooled through the region of rapid transformation so quickly that there is effectively no reaction during the cooling; the high temperature structure will then be preserved indefinitely at a sufficiently low temperature. "Quenching" to retain the high temperature structure is a standard method used in the experimental determination of metallurgical equilibrium diagrams. In general, all reactions which involve changes of composition in different regions of the specimen, i.e. long-range diffusion, will be slow, and can be prevented by rapid cooling. If no martensitic transformation intervenes, the quenching method is successful in such cases. Some reactions, for example the solidification of a liquid alloy, cannot be inhibited completely even by drastic cooling, but equilibrium can be prevented, and the non-equilibrium structure formed during the quench (which in extreme cases may be non-crystalline) is readily recognized.

(3) *Irreversibility of the transformation.* Since the individual atoms move independently, there is no correlation between the initial and final positions of the atoms after retransforming to the original phase. For example, if we convert a crystalline phase  $\alpha$  into a new structure  $\beta$  by heating through a transformation range, and then reconvert to  $\alpha$  by cooling, the assembly is thermodynamically in its original state (neglecting grain boundary and surface energies). Usually, however, the crystals of  $\alpha$  will be unrelated in size, shape, or orientation to the original  $\alpha$  crystals. The transformation is thus irreversible, not only thermodynamically, but also in the special sense that the assembly never returns to its initial configuration. In practice this may not always be obvious because nucleation may be heterogeneous and take place at certain preferred sites.

(4) *Effect of plastic deformation.* Reactions in the solid state are often accelerated by cold-working the material prior to transformation. There are several possible explanations for this effect. Nucleation may be easier in plastically deformed regions of the crystal lattice, since the effective driving force is increased and the free energy barrier decreased. Another factor is the activation energy for atomic diffusion, which may similarly be lowered in heavily deformed regions. The process of deformation leads to a large temporary increase in the number of vacant lattice sites, and this will also temporarily increase the diffusion rate, independently of the activation energy. Plastic deformation at temperatures where internal stresses can be removed rapidly by recovery or recrystallization does not affect the transformation rate.

(5) *Composition, atomic volume, and shape of the new phase.* The composition and atomic volume of the reaction products need not be related in any way to those of the original phase. Only in pure metals, polymorphic changes, order-disorder reactions, and transformations not involving thermodynamic phase changes will there be no change in composition.

In nucleation and growth transformations, the shape of the particles of the new phase when transformation is incomplete, or when the equilibrium configuration is polyphase, varies considerably. If surface energies are important, the shape will be spherical for isotropic surface energy (e.g. liquid droplets), and roughly equi-axed polyhedral for anisotropic energies. If there is a volume change, however, the strain energy in a condensed assembly will often be more important than the surface energies. The crystals of the new phase will then be in the form of plates or needles, orientated with respect to the original lattice so that the best atomic fit is obtained across the interface. The resulting microstructure is called a Widmanstätten structure.

(6) *Orientation relations.* In some nucleation and growth transformations in the solid state there is no relation between the orientations of the two lattices. When Widmanstätten structures are produced, or in the early stages of coherent growth of precipitates from solid solution, it is, however, usually found that one lattice has a fixed orientation relative to the other. Orientation relations are also commonly found when two phases are formed together, as in the solidification of a eutectic alloy or a eutectoidal reaction in the solid state.

### 3. CHARACTERISTICS OF MARTENSITIC TRANSFORMATIONS

Martensitic reactions are possible only in the solid state. They do not involve diffusion, and the composition of the product is necessarily the same as that of the original phase. In alloys which are ordered, it has been shown that the phase formed by a martensitic reaction is also ordered. Negligible mixing of the atoms thus takes place during the transformation, and it follows that the thermodynamically stable configuration of the assembly often cannot be produced by such a change.

In a martensitic reaction, a co-operative movement of many thousands of atoms occurs with a velocity approaching that of sound waves in the crystal. Thermally supplied activation energy could not account for such a process, and the concept of activation energy is not generally so useful in martensitic reactions, except in the nucleation stage. The reaction starts spontaneously at some temperature, and the parent structure is then effectively mechanically unstable. In a similar way, mechanical twinning may be regarded as a special kind of martensitic reaction in which the driving force is an externally applied stress, rather than an internal free energy difference.

The main characteristics of martensitic reactions are summarized below:

(1) *Dependence on time.* The amount of transformation is virtually independent of time. At a constant temperature, a fraction of the original phase transforms very rapidly, after which there is usually no further change. This is a primary characteristic of martensitic transformations, but in some reactions there is also a small amount of isothermal transformation to the martensite phase, and in a few cases, the change is almost completely isothermal. These differences will be discussed in detail later, but as already indicated, the isothermal characteristics are the result of thermally assisted nucleation processes.

(2) *Dependence on temperature.* The amount of transformation is characteristic of temperature, providing other variables such as grain size are held constant. The velocity of

transformation is probably independent of temperature and is usually very rapid. Transformation on cooling begins spontaneously at a fixed temperature ( $M_s$ ), and as the temperature is changed, more and more material transforms, until the temperature ( $M_f$ ) is reached, at which the change is complete. In some assemblies it is doubtful whether complete transformation ever occurs spontaneously.

At any temperature a number of single crystals of the new phase form rapidly within an original grain. On cooling to another temperature these crystals usually do not grow but new crystals are formed. In favourable circumstances, however, a single crystal of the original phase can be continuously converted to a single crystal of the martensite phase.

(3) *Reversibility of the transformation.* Martensitic reactions are very reversible in the sense that an initial atomic configuration can be repeatedly obtained. A single crystal of the original phase may, for example, transform on cooling into several crystals of the new phase. The reverse change on heating will usually result in a single crystal of the same size, shape, and orientation as the original crystal. The reversibility is associated with a temperature hysteresis and the reverse reaction begins at a temperature above  $M_s$ . Moreover, in repeated transformations, the plates (single crystals) which form on cooling have the same size and shape, and appear in the same regions of the original crystal. This behaviour probably applies in principle to all martensitic reactions. Apparent exceptions, where reversibility has not been observed, can always be attributed to interfering secondary effects: in iron-carbon alloys, for example, the martensitic phase is thermodynamically unstable and starts decomposing into stable phases (tempering) before the reverse transformation can begin.

(4) *Effect of applied stress.* Plastic deformation is much more important in martensitic reactions than in nucleation and growth changes. Application of plastic stresses at any temperature in the transformation range usually increases the amount of transformation, and the reaction can often be completed by this means. In some transformations, elastic stresses have a similar effect. When single crystals are used, the direction of the applied stress is important, and some reactions may be inhibited as well as aided by a suitably orientated stress.

Deformation above  $M_s$  may also result in the formation of the product phase, even though the temperature is too high for spontaneous reaction. The highest temperature at which martensite may be formed under stress is called  $M_d$ . In general, the reverse reaction can be aided in the same way, and a suitable stress will induce transformation below the temperature at which it begins spontaneously.

If the original phase is cold-worked in a temperature range where it is stable, e.g. at temperatures sufficiently above  $M_s$ , the resultant deformation often inhibits transformation. Providing the temperature of deformation is not high enough to permit self-annealing, the  $M_s$  temperature is depressed, and the amount of transformation found at any temperature is reduced.

(5) *Composition, atomic volume, and shape of the new phase.* In a martensitic reaction, each crystal transforms to new crystals of the same chemical composition. Volume changes are often, though not invariably, small, and in some cases are zero to within the limits of experimental error. The martensite crystals are usually flat plates, which thin towards the

extremities and so have a lenticular cross-section. There are exceptions in certain simple transformations where parallel-sided bands are formed. The plates or bands are orientated with respect to the original lattice; the plane of the lattice on which they are formed is called the habit plane. It has also been possible in some transformations, as mentioned above, to change a single crystal of the original phase into a martensite single crystal by the migration of an interface from one side of the crystal to the other. This interface lies along the habit plane.

(6) *Orientation relations.* In martensitic changes, there is always a definite relation between the orientation of the original structure and that of the new phase. As with the habit planes it is usually possible to find all crystallographically equivalent variants of the relation under suitable conditions. A single martensite plate may be a single crystal or may contain two twin orientations. In the latter case, the orientations of the twins relative to the matrix are not necessarily equivalent.

(7) *Stabilization.* We have already referred to the effects produced by cold-working the original structure. The reaction is also inhibited in another way. If the specimen is cooled to a temperature in the transformation range, held there for a period of time, and then cooled again, transformation does not begin again immediately. At all subsequent temperatures the amount of transformation is less than that produced by direct cooling to the temperature concerned. This phenomenon is referred to as stabilization. The degree of stabilization increases with the time for which the specimen is held at the temperature. Slight variations in amount of transformation with cooling velocity are also presumably to be attributed to stabilization. There is no general agreement whether or not stabilization is produced by halting the cooling above the  $M_s$  temperature.

Various attempts have been made to arrive at a satisfactory definition of a martensitic transformation. Hull (1954) and Bilby and Christian (1956) independently proposed that observation of a change of shape should be regarded as a good experimental criterion, but that is now invalidated by the existence of diffusional displacive transformations. Cohen, Olson and Clapp (1979) defined a martensitic transformation as a sub category of a range of displacive transformations. Displacive transformations dominated by atomic "shuffles" (if such exist) were first eliminated and the lattice strains were then divided into shear and dilatational components. Martensitic transformations were regarded as dominated by shear strains and only those changes in which the shear strains are larger than the dilatational strains were classified as truly martensitic. Finally, a distinction was suggested between true martensites in which the strain energy plays a major role in the kinetics, morphology etc. and quasi martensites in which the shear strains are very small so that the strain energy produced is relatively unimportant. This scheme thus defines a martensitic transformation as a "shear dominant, lattice distortive, diffusionless transformation occurring by nucleation and growth". The classification was re-examined by Christian, Olson and Cohen (1995). In the original paper, alloys such as indium-thallium were tentatively assigned to quasi martensitic status along with ferro electric and similar transitions but it is clear from experimental work that these alloys with strains of the order of 1% should be regarded as true martensites. Quasi martensites are believed to form by a continuous process (a "strain spinodal") rather than by nucleation and growth. Clapp (1995) quoted various definitions in the literature and finally recommends "the martensitic

transformation involves cooperative movement of a set of atoms across an interface causing a shape change and sound". This definition would include deformation twinning as a martensitic transformation and could encompass growth of high angle grain boundaries. It seems essential to mention the shape change as part of the definition.

#### 4. ISOTHERMAL TRANSFORMATION CURVES

In this section we give an introductory description of the theory of the relation between the fraction of the assembly transformed and the time at constant temperature. This formal theory is largely independent of the particular models used in detailed descriptions of the mechanism of transformation, and can therefore be given here before these models are discussed. The concept of nucleation rate will be given an operational definition, which may later be compared with the theoretical quantity introduced in Chapter 10.

In a homogeneous reaction, the probability of any small region transforming in a given time interval will be the same in all parts of the untransformed volume. The volume transforming in a short time interval will thus be proportional to the volume remaining untransformed at the beginning of this interval, and this leads to a first-order rate process. Suppose the total volume is  $V$  and the volume which has transformed from  $\alpha$  to  $\beta$  at any time is  $V^\beta$ . Then:

$$\frac{dV^\beta}{dt} = k(V - V^\beta) \quad \text{or} \quad V^\beta/V = 1 - \exp(-kt). \quad (4.1)$$

The constant  $k$  is called the rate constant. The rate of transformation decreases continuously with time (see Fig. 1.1).

For nucleation and growth reactions, the situation is more complex. Consider first the size of an individual transformed region. The region is formed at a time  $t = \tau$  ( $\tau$  may

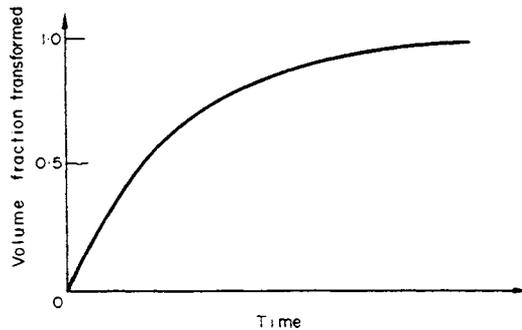


FIG. 1.1. Reaction curve for a homogeneous transformation.

be called the induction period), and thereafter its size increases continuously. If the transformation product has the same composition as the original phase, it is found experimentally that in nearly all reactions any dimension of the transformed region is a linear function of time. This is shown schematically in Fig. 1.2. The reduction of growth rate when  $t$  becomes

large is due to the mutual impingement of regions transforming from separate nuclei, which must ultimately interfere with each other's growth. The intercept on the time axis, which gives the induction period  $\tau$  for nucleation of the region, naturally cannot be observed experimentally, and is inferred by extrapolating back the linear portion of the curve. In most transformations, this procedure probably represents the actual way in which the regions form, but in some cases alternative extrapolations, such as that shown to zero time, may be more correct. Such a curve implies that stable nuclei of the new phase are already present in the assembly at the beginning of the transformation, but that the initial growth rate is an increasing function of the time.

In developing a formal theory of transformation kinetics, the distinction between these alternative physical processes leading to the formation of nuclei is irrelevant. We adopt an operational definition of the nucleation rate per unit volume,  ${}^vI$ , which is related to the reciprocal of a mean value of the period  $\tau$ . Suppose that at time  $t = \tau$  the untransformed volume is  $V^\alpha$ , and that between times  $t = \tau, \tau + \delta\tau$  a number of new regions are nucleated (i.e. reach some arbitrary minimum size). In principle, this number can be determined experimentally by plotting curves of the type shown in Fig. 1.2 for a large number of

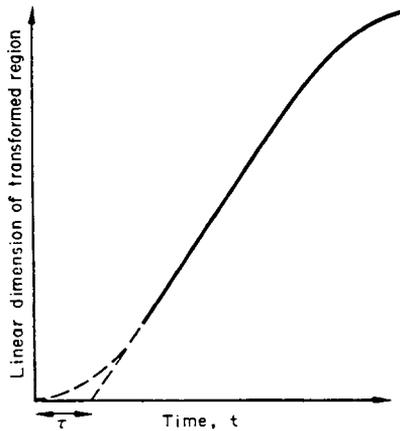


FIG. 1.2. Schematic growth curve for a product region.

regions, extrapolating back to give intercepts on the time axis, and finding the number of such intercepts between  $\tau$  and  $\tau + \delta\tau$ . This number is  ${}^vIV^\alpha \delta\tau$ , and defines  ${}^vI$  at time  $\tau$ ; in order to give statistically significant results, the number of intercepts per unit time ( ${}^vIV^\alpha$ ) must be large. The growth rate for any direction is similarly obtained by plotting the length for that direction as a function of time for a large number of regions, and finding the average slope. In general, the growth rate will be anisotropic, but in this introductory treatment, we assume an isotropic growth rate  $\Upsilon$  so that the transformed regions are spherical. This is, in fact, a good approximation in many actual changes. The volume of a  $\beta$  region originating at time  $t = \tau$  is then

$$\left. \begin{aligned} v_\tau &= (4\pi/3) \Upsilon^3(t-\tau)^3 & (t > \tau), \\ v_\tau &= 0 & (t < \tau). \end{aligned} \right\} \quad (4.2)$$

When the mean composition of the matrix must be changed during the transformation, the size of a transformed region in any direction is often found to be proportional to  $(t-\tau)^{1/2}$ , this parabolic growth law being dependent on diffusion rates. An equivalent operational definition of the nucleation rate can obviously be given, but in the present section only linear growth rates are considered.

In the whole assembly, the number of new  $\beta$  regions nucleated in the time interval between  $\tau$  and  $\tau+d\tau$  is  ${}^vIV^\alpha d\tau$ . During the initial stages of transformation, when  $V^\beta \ll V^\alpha$ , the nuclei are widely spaced, and the interference of neighbouring nuclei is negligible. Under these conditions, the transformed volume at time  $t$  resulting from regions nucleated between times  $\tau$  and  $\tau+d\tau$  is  $dV^\beta = v_\tau {}^vIV^\alpha d\tau$ . Since  $V^\alpha$  is effectively constant and equal to  $V$ , the total volume transformed at time  $t$  is thus

$$V^\beta = (4\pi V/3) \int_{\tau=0}^t {}^vI\gamma^3(t-\tau)^3 d\tau.$$

This equation can be integrated only by making some assumption about the variation of  ${}^vI$  with time. The simplest assumption is that  ${}^vI$  is constant, and this leads to the result

$$\zeta = V^\beta/V = (\pi/3) {}^vI\gamma^3 t^4, \quad (4.3)$$

where we have now introduced the symbol  $\zeta$  for the volume fraction transformed at time  $t$ . The rate of transformation according to this equation rises rapidly in the initial stages.

In a more exact treatment, we must consider the mutual interference of regions growing from separate nuclei. When two such regions impinge on each other, there are three possible consequences. The regions may unite to form a single region, as often happens with liquid droplets forming from the vapour, or they may separate, each continuing to grow (until a late stage in the transformation) as though there were no impingement. Separation can obviously occur only if the primary phase is fluid, and preferably gaseous. The third possibility is that the two regions develop a common interface, over which growth ceases, although it continues normally elsewhere. This must happen in all solid transformations, and it is the case which we shall consider. The problem is primarily geometrical, and was first treated by Kolmogorov (1937), Johnson and Mehl (1939) and Avrami (1939, 1940, 1941).

During the time  $d\tau$ , when  ${}^vIV^\alpha d\tau$  new transformed regions are nucleated, we may also consider that  ${}^vIV^\beta d\tau$  regions would have nucleated in the transformed portion of the assembly, had not transformation previously occurred there. Avrami described these as phantom nuclei, and went on to define an "extended" volume of transformed material,  $V_e^\beta$ , by the relation

$$\left. \begin{aligned} dV_e^\beta &= v_\tau {}^vI(V^\alpha + V^\beta) d\tau, \\ V_e^\beta &= (4\pi V/3) \int_0^t \gamma^3 {}^vI(t-\tau)^3 d\tau. \end{aligned} \right\} \quad (4.4)$$

i.e.

$V_e^\beta$  differs from the actual volume of transformed material in two ways. Firstly, we have counted phantom regions, nucleated in already transformed material. Secondly, we have treated all regions as though they continued growing, irrespective of other regions. The

extended volume can thus be visualized as a series of volume elements having the same limiting surface as the actual transformed volume but all growing "through" each other. Some elements of the transformed volume are counted twice, others three times, and so on, in order to obtain the extended volume. It follows that the extended volume may be larger than the real volume of the whole assembly  $V$ .

The significance of  $V_e^\beta$  is that it is simply related to the kinetic laws of growth, which may thus be separated from the geometrical problem of impingement. We have now to find a relation between  $V_e^\beta$  and  $V^\beta$ . Consider any small random region, of which a fraction  $(1 - V^\beta/V)$  remains untransformed at time  $t$ . During a further time  $dt$ , the extended volume of  $\beta$  in the region will increase by  $dV_e^\beta$ , and the true volume by  $dV^\beta$ . Of the new elements of volume which make up  $dV_e^\beta$ , a fraction  $(1 - V^\beta/V)$  on the average will lie in previously untransformed material, and thus contribute to  $dV^\beta$ , whilst the remainder of  $dV_e^\beta$  will be in already transformed material. This result clearly follows only if  $dV_e^\beta$  can be treated as a completely random volume element, and it is for this reason that phantom nuclei have to be included in the definition of  $V_e^\beta$ .

The above arguments are based on the assumption that nucleation is random, in the sense that if we divide the assembly into small equal volume elements, the probability of forming a nucleus in unit time is the same for all these elements. The treatment does not preclude the possibility that nuclei form preferentially at certain sites in the  $\beta$  phase, but as developed here, it may be applied to experimental observations only if the minimum resolvable  $\beta$  volume contains several such sites. All equal volume elements of size greater than this observational limit then have equal nucleation probabilities. In practice, nucleation may occur preferentially along macroscopic surfaces (grain boundaries) of the assembly, so that volume elements in different regions may have quite different nucleation probabilities. The formal theory can be extended to cover such cases, but discussion is postponed to Chapter 12.

We now write the relation between  $V^\beta$  and  $V_e^\beta$  in the form

$$\begin{aligned} dV^\beta &= (1 - V^\beta/V) dV_e^\beta \\ \text{or} \quad V_e^\beta &= -V \ln (1 - V^\beta/V). \end{aligned} \tag{4.5}$$

Substituting into (4.4)

$$-\ln (1 - \zeta) = (4\pi/3)\Upsilon^3 \int_0^t vI(t-\tau)^3 d\tau. \tag{4.6}$$

This equation may be integrated only by making specific assumptions about the variation of  $vI$  with time. In particular, if  $vI$  is constant

$$\zeta = 1 - \exp(-\pi\Upsilon^3 vIt^4/3). \tag{4.7}$$

Note that eqn. (4.3) is given by the first term in the expansion of the right-hand side of (4.7), so that the two expressions become identical as  $t \rightarrow 0$ , in accordance with the assumptions made in deriving (4.3).

In general,  $vI$  may not be constant, but may either increase or decrease with time. The physical processes leading to this will be described fully later, but we emphasize here that eqn.

(4.6) and not (4.7) must be used for  $\zeta$  in the general case. This is made clear by an alternative assumption, used by Avrami. He supposes that nucleation occurs only at certain preferred sites in the assembly, which are gradually exhausted. If there are  ${}^vN_0$  sites per unit volume of the  $\alpha$  phase initially, and  ${}^vN$  remaining after time  $t$ , the number disappearing in a further small time interval  $dt$  is  $d{}^vN = -{}^vN\nu_1 dt$ , where the frequency  $\nu_1$  gives the rate at which an individual site becomes a nucleus. Thus  ${}^vN = {}^vN_0 \exp(-\nu_1 t)$ , and the nucleation rate per unit volume is

$${}^vI = -d{}^vN/dt = {}^vN_0\nu_1 \exp(-\nu_1 t). \quad (4.8)$$

Substituting  $({}^vI)_{t-\tau}$  into eqn. (4.6) and integrating by parts gives

$$\zeta = 1 - \exp \left[ (8\pi {}^vN_0 \Upsilon^3 / \nu_1^3) \left\{ \exp(-\nu_1 t) - 1 + \nu_1 t - \frac{\nu_1^2 t^2}{2} + \frac{\nu_1^3 t^3}{6} \right\} \right]. \quad (4.9)$$

There are two limiting forms of this equation, corresponding to very small or very large values of  $\nu_1 t$ . Small values of  $\nu_1 t$  imply that  ${}^vI$  (eqn. (4.8)) is effectively constant, and the limiting value of (4.9), obtained by expanding  $\exp(-\nu_1 t)$ , is identical with eqn. (4.7). Large values of  $\nu_1 t$ , in contrast, mean that  ${}^vN$  quickly becomes zero, all nucleation centres being exhausted at an early stage in the reaction. The limiting value of eqn. (4.9) is then

$$\zeta = 1 - \exp\{- (4\pi {}^vN_0/3) \Upsilon^3 t^3\}. \quad (4.10)$$

Avrami proposed that for a three-dimensional nucleation and growth process, we should use the general relation

$$\zeta = 1 - \exp(-kt^n), \quad (4.11)$$

where  $3 \leq n \leq 4$ . This should cover all cases in which  ${}^vI$  is some decreasing function of time, up to the limit when  ${}^vI$  is constant.

The above treatment, whilst including the effects of impingement, neglects the effect of the free surface. Thus if transformation occurs in a thin sheet of solid material, it may happen that the average dimension of a transformed region is much greater than the thickness of the sheet. Growth in this direction must soon cease, and thereafter the growth is essentially two-dimensional. Instead of (4.2), the expression for the volume of an individual transformed region becomes

$$v'_\tau = \pi \delta \Upsilon^2 (t - \tau)^2, \quad (4.12)$$

where  $\delta$  is the sheet thickness. Similarly, for a wire of diameter  $\delta$ , the growth would be one-dimensional, the volume of each transformed region being

$$v''_\tau = (\pi/4) \delta^2 \Upsilon (t - \tau). \quad (4.13)$$

The use of  $v'$  or  $v''$  in the definition of the extended volume (eqn. (4.4)) will introduce corresponding modifications in the expressions for  $\zeta$ . In the Avrami theory of nucleation at preferred sites, it is readily seen that the general expression for the volume transformed (eqn. (4.11)) remains valid if  $2 \leq n \leq 3$  (two-dimensional growth) and  $1 \leq n \leq 2$  (one-dimensional growth).

The isothermal transformation curves obtained by substituting  $n = 1$  in eqn. (4.11) is equivalent to that for a first-order homogeneous reaction. All the other possibilities, however, give sigmoidal curves for  $\zeta$  against  $t$ , in which the fractional volume transformed increases slowly at first, then much more rapidly, and, finally, slowly again. Most experimental transformation curves are sigmoidal in shape; a typical example is shown in Fig. 1.3. Avrami pointed out that if we plot  $\zeta$  against  $\ln t$ , all the curves with the same value of  $n$  will have the same shape, and will differ only in the value of  $k$ , which is equivalent to a change of scale. He therefore proposed that the "shape" of a reaction curve should be defined by a  $\zeta - \ln t$  plot. A more useful plot is  $\log \log [1/(1-\zeta)]$  against  $\log t$ , the slope of which gives  $n$ .

In the general form of eqn. (4.11), the above theory applies to many real transformations. However, it cannot be safely applied in practice unless the assumptions about  $\nu I$  and  $Y$

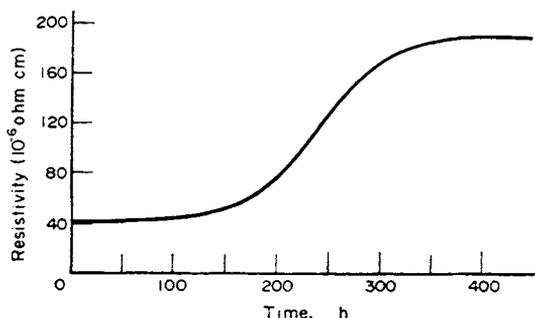


FIG. 1.3. Kinetics of transformation  $\beta \rightarrow \alpha$  manganese at 25°C, as shown by electrical resistivity measurements (after Potter *et al.*, 1949).

can be verified. The Avrami theory assumes that the nucleation frequency is either constant, or else is a maximum at the beginning of transformation and decreases (slowly or rapidly) during the course of transformation. In contrast to this, there may be an operational rate of nucleation which increases with time. The Avrami equation is also often applied to transformations in which the growth rate is diffusion controlled, but in most cases there is no adequate theoretical sanction for this. These questions are fully discussed in Chapter 12, which takes up the formal theory of this section in greater detail.

#### REFERENCES

- AVRAMI, M. (1939) *J. Chem. Phys.* **7**, 1103; (1940) *Ibid.* **8**, 212; (1941) *Ibid.* **9**, 177.  
 BILBY, B.A. and CHRISTIAN, J.W. (1956) *The Mechanism of Phase Transformations in Metals*, p.121, The Institute of Metals, London.  
 BUERGER, M. J. (1951) *Phase Transformations in Solids*, p. 183, Wiley, New York.  
 CHRISTIAN, J. W. (1965) *Physical Properties of Martensite and Bainite*, Iron and Steel Institute Spec. Rep. **93**, 1; (1997) *Progr. in Mat. Sci.* **42**, 101.  
 CHRISTIAN, J. W., OLSON, G.B. and COHEN, M. (1995) *J. de Phys.* **C 3-3**.  
 CLAPP, P.C. (1995) *J. de Phys.* **C 3-11**.

- CLARK, H. M. and WAYMAN, C. M. (1970) *Phase Transformations*, p. 59, American Society for Metals, Metals Park, Ohio.
- COHEN, M., CAHN, J. W., CHRISTIAN, J. W., FLINN, P. A., HILLERT, M., KAUFMAN, L., and READ, T. A. (1965) *Perspectives in Materials Research*, p. 309, Office of Naval Research, Washington.
- COHEN, M., OLSON, G.B. and CLAPP, P.C. (1979) *Proc. ICOMAT - 79*, p. 10, M.I.T. Press, Cambridge, Mass.
- DAW, M.S., and BASKES, M.I. (1984) *Phys. Rev. D* **29**, 6448.
- FINNIS, M.W. and SINCLAIR, J.E. (1984) *Phil. Mag.* **A 50**, 45.
- FRANK, F. C. (1963) *Relation between Structure and Strength in Metals and Alloys*, p. 248, HMSO, London.
- GUGGENHEIM, E. A. (1967) *Thermodynamics*, 5th ed., North-Holland, Amsterdam.
- GUY, A. G. (1972) *Met. Trans.* **3**, 2535.
- HOHENBERG, P. and KOHN, W. (1964) *Phys. Rev. B* **136**, 264.
- HULL, D. (1954) *Bull. Inst. Metals* **2**, 134.
- JOHNSON, W. A. and MEHL, R. F. (1939) *Trans. Am. Inst. Min. (Metall.) Engrs.* **135**, 416.
- KOHN, W. and SHAM, L.J. (1965) *Phys. Rev. A* **149**, 1133.
- KOLMOGOROV, A.N. (1937) *Bull. Acad. Sci. USSR Phys. Ser.* **1**, 355.
- LIEBERMAN, D. S. (1969) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 167, Institute of Metals, London.
- LIEBERMAN, D. S. (1970) *Phase Transformations*, p. 1, American Society for Metals, Metals Park, Ohio.
- PETTIFOR, D.G. (1989) *Phys. Rev. Letters* **63**, 2480.
- PETTIFOR, D.G. and AOKI, M. (1991) *Phil. Trans. Roy. Soc. A* **334**, 439.
- POTTER, E. V., LUKENS, H. C., and HUBER, R. W. (1949) *Trans. Am. Inst. Min. (Metall.) Engrs.* **185**, 399.
- SUTTON, A.P. and BALLUFFI, R.W. (1995) *Interfaces in Crystalline Materials*, The Clarendon press, Oxford.
- SUTTON, A.P., FINNIS, M.W., PETTIFOR, D.G. and OHTA, Y. (1988) *J. Phys. C (Solid State Phys.)* **21**, 33.
- TERSOFF, O. (1988) *Phys. Rev. B* **38**, 9002.

## CHAPTER 2

### *Formal Geometry of Crystal Lattices*

#### 5. DESCRIPTION OF THE IDEAL CRYSTAL

Amorphous and quasi-crystalline forms of solid have been much investigated in recent years, but most solid metals are crystalline and some appreciation of crystallography is essential to a study of metallic transformations. The scientific concept of a crystal has evolved gradually from the original classification by external shape to modern views on the internal atomic arrangement. The recognition that the distinguishing feature of crystalline solids lies in their regular internal arrangement led to a description which Zachariasen (1945) has termed the macroscopic concept of a crystal. The macroscopic crystal is defined in terms of physical properties which have precise meaning (or, at least, are measurable) only for regions containing appreciable numbers of atoms. Such properties are invariant with respect to a translation within an infinite crystal, but (except for scalar properties) not with respect to a rotation. A crystal is thus a homogeneous, anisotropic solid; a noncrystalline, or amorphous, solid is both homogeneous and isotropic.

The development of X-ray methods enabled the structure of a crystal to be investigated on a finer scale. It was then found that crystals are not truly homogeneous, but the arrangement of atoms is periodic in three dimensions. This is the familiar modern picture, which we shall take as our starting point, but we emphasize here that it is only an abstraction from the much less ordered situation in a real crystal. In recent years, attention has been directed especially to the imperfections in real crystal structures. These imperfections represent comparatively small deviations from the mathematical concept of an ideal crystal, but they nevertheless control many of the most important physical properties. In this chapter we are concerned only with the ideal crystal; the nature of the approximations involved in this description, and the extent to which a real crystal may be considered to be an ideal crystal containing imperfections, will be considered in detail later.

The ideal crystal may be regarded as the repetition in three dimensions of some unit of structure, within which the position of each atom is specified exactly by a set of spatial coordinates. Let us choose an origin within the crystal. This will be one of an infinite set of points, each possessing an identical configuration of surrounding atoms. The positions of these points may be represented by the vectors

$$\mathbf{u} = u_i \mathbf{a}_i \quad (i = 1, 2, 3), \quad (5.1)$$

where  $u_i$  have only integral values, and the summation convention is used. The translation between any two lattice points is a lattice vector, and the three non-coplanar lattice vectors

$\mathbf{a}_i$  outline a parallelepiped known as the unit cell. Any other set of lattice vectors, formed from linear combinations of the set  $\mathbf{a}_i$ , may also be used as basic vectors, so there is an infinite number of possible unit cells. The volume of the unit cell is given by the scalar triple product of the vectors which outline it; when this volume is a minimum the vectors are primitive vectors, and define a unit cell of the Bravais lattice. Each such primitive cell contains only one lattice point; cells of larger volume contain two or more lattice points. The primitive unit cell may still be chosen in an infinity of ways, since we have placed no restriction on its shape. The most useful unit cell is usually that in which the vectors  $\mathbf{a}_i$  are as nearly as possible of equal scalar magnitude.

The quantities  $u_i$  of eqn. (5.1) give the position of the end point of  $\mathbf{u}$  in an oblique Cartesian coordinate system, in which distances along the axes are measured in multiples of the lengths of the basic vectors  $\mathbf{a}_i$ . Such a coordinate system forms a natural framework for representation of the crystal lattice, and is generally preferable to the alternative method of using coordinate axes parallel to the vectors  $\mathbf{a}_i$ , but having unit measure lengths. In the latter system, the coordinates of a lattice point are  $u_{(i)} | \mathbf{a}_{(i)}$  (the brackets indicating suspension of the summation rule), and these are sometimes called the physical components of the vector  $\mathbf{u}$ . However, it is often convenient to use a coordinate system with orthogonal axes of equal base lengths, and this is called an orthonormal system. It is defined by the three unit vectors  $\mathbf{i}_i$ , which satisfy the relations

$$\mathbf{i}_i \cdot \mathbf{i}_j = \delta_{ij}, \quad (5.2)$$

where  $\delta_{ij}$ , called the Kronecker delta, is equal to unity when  $i = j$ , and to zero when  $i \neq j$ . In terms of the orthonormal system, a vector  $\mathbf{x}$  may be written as

$$\mathbf{x} = x_i \mathbf{i}_i. \quad (5.3)$$

In tensor analysis or matrix algebra,<sup>†</sup> the vector  $\mathbf{u}$  is regarded as the array of numbers  $u_i$ , which form its components. We may, for example, write  $\mathbf{u}$  as a single row matrix ( $u_1 \ u_2 \ u_3$ )

or as a single column matrix  $\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ , and these arrays would conventionally be given the same

symbol as  $\mathbf{u}$ . This sometimes leads to confusion, and it is desirable to have a way of distinguishing between the vector  $\mathbf{u}$ , which is a physical entity, and its matrix representation, which depends on a particular coordinate system. We shall do this by using a notation in which **bold face type** is used for symbols representing physical quantities (vectors and tensors), and **sans serif type** is used for their matrix representations. In addition, we specify that

column matrices may be written  $\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = [u_1 \ u_2 \ u_3]$  for convenience, whilst row matrices

<sup>†</sup> The formal theory of this chapter may be expressed in either tensor or matrix notation. We find it more convenient to use the latter, but many of the equations will be given in both forms to facilitate reference to other books and papers. A brief description of the more important features of tensor notation is given on pp. 35-8.

are written  $(u_1 u_2 u_3)$ . We thus have

$$\mathbf{u} = [u_1 u_2 u_3] \quad \mathbf{x} = [x_1 x_2 x_3], \quad (5.4)$$

where  $\mathbf{u}$  and  $\mathbf{x}$  are representations of the vectors  $\mathbf{u}$  and  $\mathbf{x}$ . Wherever possible we shall keep to the practice that a letter such as  $\mathbf{u}$  (and later  $\mathbf{v}$ ) will represent a vector which is most conveniently expressed in a lattice coordinate system, while a letter such as  $\mathbf{x}$  (and later  $\mathbf{y}$ ) will represent a vector which is usually referred to an orthonormal system.

Any matrix of  $n$  rows and  $m$  columns may be formed into a new matrix of  $m$  rows and  $n$  columns (the *transposed* matrix) by interchanging the rows and columns. Clearly, the single row and single column matrix representations of  $\mathbf{u}$  are the transposes of each other, so that eqn. (5.4) also implies the notation

$$\mathbf{u}' = (u_1 u_2 u_3) \quad \mathbf{x}' = (x_1 x_2 x_3), \quad (5.5)$$

where  $\mathbf{u}'$  is the symbol for the transposed matrix formed from  $\mathbf{u}$ .

Later in this chapter we need to refer the vector  $\mathbf{u}$  to other sets of coordinate axes. We define a second set of axes by the base vectors  $\mathbf{b}_i$ , and the vector  $\mathbf{u}$  then has a different matrix representation. We may distinguish between the two representations when required by describing the set of vectors  $\mathbf{a}_i$  as the basis  $\mathbf{A}$ , and the set of vectors  $\mathbf{b}_i$  as the basis  $\mathbf{B}$ . The column matrix representations are then written as  ${}^A\mathbf{u} = [{}^A u_1 {}^A u_2 {}^A u_3]$  and  ${}^B\mathbf{u} = [{}^B u_1 {}^B u_2 {}^B u_3]$ . Symbols of this type are sufficient for most purposes, but occasionally extra clarity is achieved by use of an extended notation, such as that used by Bowles and Mackenzie (1954). In the extended notation, the representations of  $\mathbf{u}$  as column matrices are

$${}^A\mathbf{u} = [\mathbf{A}; \mathbf{u}] \quad {}^B\mathbf{u} = [\mathbf{B}; \mathbf{u}], \quad (5.6)$$

and the corresponding row matrices are

$${}^A\mathbf{u}' = (\mathbf{u}; \mathbf{A}) \quad {}^B\mathbf{u}' = (\mathbf{u}; \mathbf{B}) \quad (5.7)$$

in which both the round brackets and the reversal of the order of the vector symbol  $\mathbf{u}$  and the base symbol  $\mathbf{A}$  or  $\mathbf{B}$  signify the transposition of the column matrices  ${}^A\mathbf{u}$ ,  ${}^B\mathbf{u}$ . The same notation is applied to vectors referred to an orthonormal basis, for which we shall usually use the symbol  $\mathbf{I}$ . Hence two different vectors  $\mathbf{u}$  and  $\mathbf{x}$  have representations

$${}^I\mathbf{u} = [\mathbf{I}; \mathbf{u}] \quad {}^I\mathbf{x} = [\mathbf{I}; \mathbf{x}]$$

in such a basis.

The advantages of the extended notation will become apparent later, but it is often sufficient to use the sans serif symbols. When no confusion about the bases is possible, the identifying superscripts will be omitted, as in eqn. (5.4).

A clear distinction must be made between the lattice points of the unit cell, and the positions of the atoms within the cell. The simplest types of crystal structure are obtained by placing an atom at each point of the lattice, and this category includes two of the three common metallic structures. More generally, the primitive Bravais lattice only gives the interval over which the unit of pattern, or motif unit, is repeated. This unit may be a

single atom or a more complex atom group; in the latter case, the structure is said to have a basis. The repeating properties of the lattice then require that if an atom of kind  $A$  is situated in a given position with respect to the origin, a similar atom is similarly situated with respect to each of the lattice points.<sup>†</sup> If there are  $r$  atoms of this kind within the unit cell, their positions with respect to the origin of the cell may be specified by the vectors  $\xi_{A,1}$ ,  $\xi_{A,2}$ , ...,  $\xi_{A,r}$ , and hence the positions of all the atoms of type  $A$  are given by

$$\mathbf{u}_{A,n} = \mathbf{u} + \xi_{A,n} \quad (n = 1, 2, \dots, r).$$

Often we do not need to specify the type of atom, so that the subscript  $A$  may be omitted and the equation may be written

$$\mathbf{u}_n = \mathbf{u} + \xi_n = (u_i + \xi_{n,i})\mathbf{a}_i \quad (n = 1, 2, \dots, r). \quad (5.8)$$

Whilst the components  $u_i$  are all integers, the components  $\xi_{n,i}$  are all less than unity. For a realistic choice of motif unit, the restriction  $|\xi_{n,i}| \leq \frac{1}{2}$  will usually be valid. When the atomic arrangement is centrosymmetric, it is possible to choose an origin so that for each atom in a position  $\xi$ , there is a similar atom in a position  $-\xi$ .

A structure which contains  $r$  atoms in the unit cell may be discussed in two different ways. We may think of a single lattice framework, at each point of which is situated a motif unit of  $r$  atoms, or we may consider the whole structure to be composed of  $r$  interpenetrating simple Bravais lattices. In many inorganic and organic crystals, the motif units have some physical significance, since they are the molecules of the compound. This is not true for most metallic structures, and it is sometimes possible to choose motif units in various different ways, all having equal validity. The alternative description may then be useful, and we shall write of single, double, etc., lattice structures, meaning structures in which the primitive unit cell contains one, two, etc., atoms. All single and double lattice structures are centrosymmetric, and the structures of most metals fall into one of these two groups.

The ideal crystal is classified by considering the symmetry properties of the atomic arrangement. There are 230 space groups, or combinations of symmetry elements, but most of these are obtained from relations between the vectors  $\xi_{A,n}$ ,  $\xi_{B,n}$ , etc. The symmetry properties of the lattice are much more restricted, and there are only fourteen Bravais lattices, obtained from relations between the vectors  $\mathbf{a}_i$ . Instead of the primitive unit cell, it is often convenient to use a larger unit cell which illustrates the symmetry of the lattice positions. For example, if the unit cell of the Bravais lattice has rhombohedral shape, and the angles between the axes are either  $109^\circ 30'$  or  $60^\circ$ , it is readily shown that the lattice positions have cubic symmetry. The conventional unit cells are cubic, and contain two and four points of the Bravais lattices respectively; the lattices are called body-centred and face-centred cubic. If we place an atom on each point of these lattices, we obtain the two common cubic metallic structures.

<sup>†</sup> This statement has to be modified in the case of a substitutional solid solution in which the possible atomic positions are occupied more or less randomly. The structure is then periodic only if the differences between the atoms of different species are ignored, so that for these atoms,  $\xi_A = \xi_B =$ , etc.

The set of integers  $u_i$  defines a translation from the origin to some lattice point. An alternative set of integers,  $U_i = C u_i$  gives a parallel translation to a lattice point which is  $C$  times more distant from the origin. If we assume that the set  $u_i$  contains no common factor, we may use the equation

$$\mathbf{u} = C_1 u_i \mathbf{a}_i = U_i \mathbf{a}_i \tag{5.9}$$

in which  $C_1$  takes all possible integral values from  $-\infty$  to  $+\infty$  to represent all lattice points in a straight line passing through the origin. In a similar way, the equation

$$\mathbf{u} = \mathbf{u}_k + U_i \mathbf{a}_i \tag{5.10}$$

represents all lattice points in a parallel straight line passing through a lattice point with position vector  $\mathbf{u}_k$ . For any given set of rational values  $u_i$ , eqn. (5.10) represents all the lattice points. The lattice structure may thus be regarded as rows of points on parallel straight lines. In conventional crystallographic notation, the quantities  $u_i$  are called the direction indices of the line, and are enclosed in square brackets  $[u_1 u_2 u_3]$ . The direction indices of a line are thus given by its representation as a column matrix  $\mathbf{u}$ . When the symmetry of the lattice requires that certain directions are equivalent, the whole set of such directions may be represented by the symbol  $\langle u_1 u_2 u_3 \rangle$ . When  $u_i$  have non-integral values, the direction they specify does not lie along a row of lattice points, and is called irrational.

Since the lattice points are arranged along straight lines, it is also possible to regard them as situated on planes. Consider first the plane defined by the basic vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . The normal to this plane is parallel to the vector  $\mathbf{a}_1 \wedge \mathbf{a}_2$ , and may be denoted by a vector  $\mathbf{a}_3^* = (\mathbf{a}_1 \wedge \mathbf{a}_2) / v_a$ , where  $v_a$  is a scalar constant. The area of the face of the unit cell formed by the vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$  is numerically equal to the length of the vector  $\mathbf{a}_1 \wedge \mathbf{a}_2$ , and since the volume of the unit cell is given by the scalar triple product  $(\mathbf{a}_1 \mathbf{a}_2 \mathbf{a}_3) = \mathbf{a}_3 \cdot \mathbf{a}_1 \wedge \mathbf{a}_2$ , we have the distance between adjacent lattice planes in the  $\mathbf{a}_3^*$  direction is equal to  $\mathbf{a}_3 \cdot \mathbf{a}_1 \wedge \mathbf{a}_2 / |\mathbf{a}_1 \wedge \mathbf{a}_2|$ . For reasons which will soon be evident, it is convenient to take the constant  $v_a$  equal to the volume of the unit cell, so that  $|\mathbf{a}_3^*|$  is equal to the reciprocal of the interplanar spacing. We have then  $\mathbf{a}_3 \cdot \mathbf{a}_3^* = 1$ . Proceeding in the same way for the other two faces of the unit cell, we obtain a set of vectors  $\mathbf{a}_i^*$  perpendicular to the faces of the cell and satisfying the relations:

$$\mathbf{a}_i \cdot \mathbf{a}_j^* = \delta_{ij}. \tag{5.11}$$

The vectors  $\mathbf{a}_i^*$  are said to be reciprocal to the vectors  $\mathbf{a}_i$ . We can now exactly reverse the above reasoning by writing  $\mathbf{a}_1 = \mathbf{a}_2^* \wedge \mathbf{a}_3^* / v_a^*$ , and since  $\mathbf{a}_1 \cdot \mathbf{a}_1^* = 1$ , it follows that  $v_a^* = (\mathbf{a}_1^* \mathbf{a}_2^* \mathbf{a}_3^*)$ , the volume of the parallelepiped formed by the reciprocal set  $\mathbf{a}_i^*$ . The relations between the two sets of vectors are thus symmetrical, and each is reciprocal to the other. The two volumes are also reciprocal and

$$v_a v_a^* = 1. \tag{5.12}$$

The set of reciprocal vectors,  $\mathbf{a}_i^*$ , may be used to define a new coordinate system, which we describe as basis  $A^*$ , and we then have

$$\mathbf{u} = u_i^* \mathbf{a}_i^*,$$

so that  $\mathbf{u}$  is represented by<sup>†</sup>

$$\mathbf{u}^* = [u_1^* \ u_2^* \ u_3^*] = [A^*; \ \mathbf{u}].$$

Suppose we form the scalar product of  $\mathbf{u}$  and one of the reciprocal vectors  $\mathbf{a}_i^*$ . This gives

$$\mathbf{u} \cdot \mathbf{a}_i^* = u_j \mathbf{a}_j \cdot \mathbf{a}_i^* = u_i \tag{5.13}$$

and, similarly,

$$\mathbf{u} \cdot \mathbf{a}_i = u_i^*. \tag{5.14}$$

These relations are shown in two dimensional form in Fig. 2.1. In the axis system  $A$ , the components  $u_i$ , measured in units of  $|\mathbf{a}_i|$  give the displacements parallel to the axes which

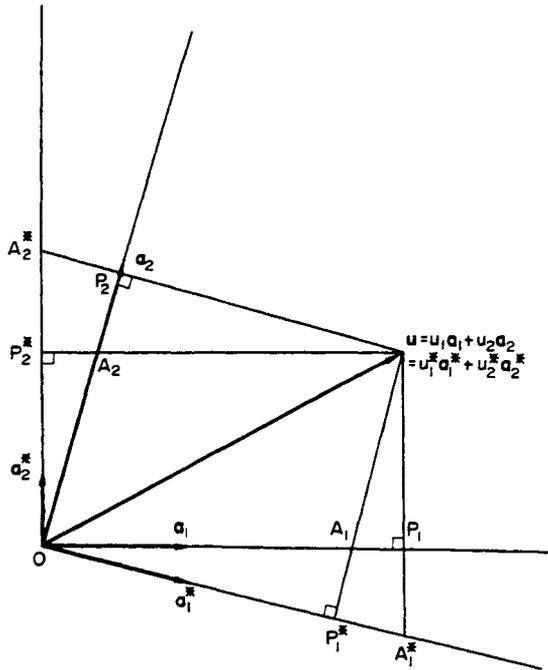


FIG. 2.1. Covariant and contravariant components of  $\mathbf{u}$ .

$OA_1 = u_1 a_1$	$OA_1^* = u_1^* a_1^*$	$OP_1 = u_1^* / a_1$	$OP_1^* = u_1 / a_1^*$
$OA_2 = u_2 a_2$	$OA_2^* = u_2^* a_2^*$	$OP_2 = u_2^* / a_2$	$OP_2^* = u_2 / a_2^*$

add together to give  $\mathbf{u}$ . The components  $u_i^*$  in the same basis are the projections of  $\mathbf{u}$  along the axes, measured in units of  $1/|\mathbf{a}_i|$ . In tensor analysis (see p. 35), the quantities  $u_i$  are referred to as the contravariant components of  $\mathbf{u}$ , and  $u_i^*$  are the covariant components of  $\mathbf{u}$ . In the basis  $A^*$ , the interpretations of  $u_i$  and  $u_i^*$  are reversed, as shown in Fig. 2.1.

<sup>†</sup> In matrix algebra, the notation  $\mathbf{u}^*$  commonly means the matrix which is the complex conjugate of the matrix  $\mathbf{u}$ . In this book, all matrices are real, so no confusion arises.

We can now write useful expressions for the scalar and vector product of two vectors  $\mathbf{u}$  and  $\mathbf{v}$ . For the scalar product

$$\mathbf{u} \cdot \mathbf{v} = (u_i \mathbf{a}_i) \cdot (v_j^* \mathbf{a}_j^*) = u_i v_i^* = u_i^* v_i. \quad (5.15)$$

In matrix form this equation is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}' \mathbf{v}^* = \mathbf{v}' \mathbf{u}^*$$

and, in particular, the length of the vector  $\mathbf{u}$  is given by

$$|\mathbf{u}|^2 = \mathbf{u}' \mathbf{u}^*. \quad (5.16)$$

Equation (5.16) involves both sets of components  $u_i$  and  $u_i^*$ , but it is obvious that the length of the vector can be expressed in terms of the components  $u_i$  alone. In order to do this, we must find the relation between the bases  $\Lambda$  and  $\Lambda^*$ , and this is discussed in the next section.

For the vector product of  $\mathbf{u}$  and  $\mathbf{v}$ , we have

$$\mathbf{u} \wedge \mathbf{v} = (u_j \mathbf{a}_j) \wedge (v_k \mathbf{a}_k) = v_a \varepsilon_{ijk} u_j v_k \mathbf{a}_i^*, \quad (5.17)$$

where  $\varepsilon_{ijk} = 0$  unless  $i, j, k$  are all different, and  $\varepsilon_{ijk} = \pm 1$  according to whether  $i, j, k$  are even or odd permutations of 1, 2, 3. If we write  $\mathbf{w} = (\mathbf{u} \wedge \mathbf{v})$ , eqn. (5.17) gives the components of  $\mathbf{w}$  in the basis  $\Lambda^*$  as

$$w_i^* = v_a \varepsilon_{ijk} u_j v_k. \quad (5.18)$$

Clearly the components of  $\mathbf{w}$  in the basis  $\Lambda$  are given by

$$w_i = v_a^* \varepsilon_{ijk} u_j^* v_k^*.$$

The orientation of any plane is completely specified by the relative lengths of the intercepts it makes on the axes  $\mathbf{a}_i$ . Consider a plane which intersects the axes at points distant  $h_1^{-1}$ ,  $h_2^{-1}$ ,  $h_3^{-1}$  from the origin. This plane contains the two vectors  $(\mathbf{a}_1/h_1 - \mathbf{a}_2/h_2)$  and  $(\mathbf{a}_2/h_2 - \mathbf{a}_3/h_3)$ , and its normal is then parallel to their vector product, i.e. to a vector

$$\mathbf{Ch} = \frac{(\mathbf{a}_1 \mathbf{a}_2 \mathbf{a}_3)}{h_1 h_2 h_3} (h_1 \mathbf{a}_1^* + h_2 \mathbf{a}_2^* + h_3 \mathbf{a}_3^*).$$

Leaving out the scalar multiplying factor, we may use the normal vector

$$\mathbf{h} = h_i \mathbf{a}_i^* \quad (5.19)$$

to represent the plane. The numbers  $h_i$  are the components of the vector  $\mathbf{h}$  in the basis  $\Lambda^*$ , and for consistency of notation we should have written them  $h_i^*$ . The matrix representation of  $\mathbf{h}$  is then

$$\mathbf{h}^* = [h_1^* \ h_2^* \ h_3^*] = [\Lambda^*; \mathbf{h}].$$

Since we shall always refer vectors which represent plane normals to the reciprocal basis  $\Lambda^*$ , we may omit the asterisks and use  $h_i$  for these components. Note that the magnitude of  $\mathbf{h}$  has, at present, been chosen arbitrarily.

The triad of numbers  $h_i$  represent the orientations of all parallel planes which intersect the axes at distances  $C_2/h_i$ . The scalar equation of these planes is

$$h_1u_1 + h_2u_2 + h_3u_3 - C_2 = 0, \quad (5.20)$$

where each  $u_i$  is now regarded as a continuous variable. We are interested in those planes which contain sets of equivalent lattice points. A plane containing three lattice points includes two lattice directions, and hence an infinite set of points. The coordinates of any three points are  $[U_1 U_2 U_3]$ ,  $[V_1 V_2 V_3]$ , and  $[W_1 W_2 W_3]$ . From eqn. (5.20) and the three equations obtained by substituting these values into it, we obtain the determinantal equation

$$\begin{vmatrix} u_1 & u_2 & u_3 & -1 \\ U_1 & U_2 & U_3 & -1 \\ V_1 & V_2 & V_3 & -1 \\ W_1 & W_2 & W_3 & -1 \end{vmatrix} = 0, \quad (5.21)$$

and since the quantities  $U_i, V_i, W_i$  are all integers, it follows from (5.20) and (5.21) that  $h_1, h_2, h_3$  and  $C_2$  are also integral. Planes in which  $h_i$  are not integral do not contain lattice points, and are termed irrational. Suppose first that  $h_i$  contains no common factor. For each integral value of  $C_2$  between  $-\infty$  and  $+\infty$ , there is an infinite number of points  $u_i = U_i$  satisfying eqn. (5.20). As  $C_2$  varies, (5.20) thus represents an infinite set of equally spaced parallel planes, containing all the lattice points. One of these planes ( $C_2 = 0$ ) passes through the origin, and the next plane of the set has  $C_2 = 1$ . As already shown, the vector  $\mathbf{h}$  is perpendicular to the set of planes, and the interplanar spacing is thus given by the projection of the vector  $\mathbf{a}_1/h_1$  on the direction of  $\mathbf{h}$ , i.e. by the relation

$$d_h = \frac{\mathbf{h} \cdot \mathbf{a}_1 / h_1}{|\mathbf{h}|} = \frac{1}{|\mathbf{h}|}. \quad (5.22)$$

The vectors  $\mathbf{h}$  are thus not only perpendicular to the planes having indices  $h_i$ , but are of length equal to the reciprocal of the interplanar spacing. The orientation of a plane is usually specified in crystallography by placing the indices  $h_i$  in round brackets,  $(h_1 h_2 h_3)$ , and this is equivalent to writing the vector as a row matrix  $\mathbf{h}' = (\mathbf{h}; \mathbf{A}^*)$ . When there are several equivalent planes, they are indicated by the symbol  $\{h_1 h_2 h_3\}$ .

We now consider the set of equations obtained by replacing the  $h_i$  in eqn. (5.19) by quantities  $H_i = C_3 h_i$  ( $C_3$  integral), it being assumed as before that the set  $h_i$  contains no common factor. As  $C_3$  varies from  $-\infty$  to  $+\infty$ , we again obtain an infinite set of planes parallel to the first set but spaced  $C_3$  times more closely. Equation (5.20) can now be satisfied with quantities  $u_i = U_i$  only if  $C_2$  is an integral multiple of  $C_3$ . When this happens, the plane is one of the original set  $h_i$ . We thus see that the vector:

$$\mathbf{h} = C_3 h_i \mathbf{a}_i^* = H_i \mathbf{a}_i^* \quad (5.23)$$

represents a set of parallel planes of spacing,  $1/|\mathbf{h}|$ , but that only every  $C_3$ th plane of the set passes through equivalent lattice points.

Equation (5.23) shows that the vectors  $\mathbf{h}$  define a lattice, and this is called the reciprocal lattice. Each point of the reciprocal lattice corresponds to an infinite set of parallel crystal planes; similarly, each set of reciprocal lattice planes is associated with a point of the real lattice. The linear vector space which is defined by the vectors  $\mathbf{a}_i^*$  is called reciprocal space. The concepts of reciprocal space, and of the reciprocal lattice, have proved very useful in crystal geometry and in the theory of X-ray diffraction.

## 6. LINEAR TRANSFORMATIONS OF THE COORDINATE SYSTEM

The position of any lattice point is given by its coordinates  $u_i$  with respect to some chosen origin and set of base vectors  $\mathbf{A}$ . As previously emphasized, the choice of basis is arbitrary, and it is often desirable to use a new basis  $\mathbf{B}$  defined by the vectors  $\mathbf{b}_i$ . We then have to develop the appropriate transformation formulae connecting  ${}^A u_i$  and  ${}^B u_i$  for any direction, and  ${}^A h_i$  and  ${}^B h_i$  for any plane.

The new basic vector  $\mathbf{b}_1$  will differ from  $\mathbf{a}_1$  both in magnitude and direction. Since any four vectors are linearly dependent, however, we can write  $\mathbf{a}_1$  as a linear function of the new vectors  $\mathbf{b}_i$ :

$$\mathbf{a}_1 = J_{11}\mathbf{b}_1 + J_{21}\mathbf{b}_2 + J_{31}\mathbf{b}_3.$$

This relation signifies that the direction indices of the vector  $\mathbf{a}_1$  referred to the basis  $\mathbf{B}$  are  $[J_{11} J_{21} J_{31}]$ .<sup>†</sup> The transformation from one basis to the other may thus be represented completely by three equations of the above form, or in the usual summation convention

$$\mathbf{a}_i = J_{ji}\mathbf{b}_j. \quad (6.1)$$

Note that although written in subscript form, this is a vector equation and not a relation between vector components. When we deal with base vectors, the subscript identifies a particular vector of the set  $\mathbf{A}$  or  $\mathbf{B}$ , rather than a particular component. The vector  $\mathbf{a}_i$  may be represented in its own basis system  $\mathbf{A}$  by the column matrix  $\mathbf{a}_i$ ; obviously the  $k$ th component of  $\mathbf{a}_i$  in this representation is  $(a_i)_k = \delta_{ik}$ .

If we now write the set of vectors  $\mathbf{a}_i$  which constitute the basis  $\mathbf{A}$  as a column matrix of vectors,  $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \mathbf{a}_3]$ , and the set  $\mathbf{b}_i$  as a similar matrix  $\mathbf{B}$ , the three equations may be combined in the matrix equation

$$\mathbf{A} = \mathbf{J}' \mathbf{B}. \quad (6.2)$$

Here  $\mathbf{J}$  is the  $3 \times 3$  array with elements  $J_{ij}$ , and the transposed matrix  $\mathbf{J}'$  is the corresponding array with elements  $J_{ji}$ .  $\mathbf{A}$  is obtained by multiplying  $\mathbf{J}'$  and  $\mathbf{B}$  in accordance with the laws of matrix algebra.

The components of each individual vector of the set  $\mathbf{a}_i$  are given by the representations

$$[\mathbf{A}; \mathbf{a}_i] = [\delta_{1i} \delta_{2i} \delta_{3i}] \quad \text{and} \quad [\mathbf{B}; \mathbf{a}_i] = [J_{1i} J_{2i} J_{3i}].$$

<sup>†</sup> We write the direction indices of  $\mathbf{a}_1$  in this form, rather than as  $[J_{11} J_{12} J_{13}]$  in order that the matrix array  $\mathbf{J} = J_{ij}$  introduced below shall give directly the transformation of vector *components*.

By comparing coefficients, it is readily seen that these quantities are related by the matrix equation

$$[\mathbf{B}; \mathbf{a}_i] = \mathbf{J}[\mathbf{A}; \mathbf{a}_i]. \quad (6.3)$$

We shall now show that this is a particular example of the general equation relating the components  ${}^A u_i$ ,  ${}^B u_i$  of any vector  $\mathbf{u}$ . Writing  $\mathbf{u} = {}^A u_i \mathbf{a}_i = {}^B u_i \mathbf{b}_i$ , and substituting for  $\mathbf{a}_i$  from (6.1),

$${}^B u_i \mathbf{b}_i = {}^A u_j J_{ji} \mathbf{b}_j = {}^A u_j J_{ij} \mathbf{b}_i$$

since both  $i$  and  $j$  are dummy indices (to be summed) on the right-hand side. The expression is now an identity, and the coefficients of corresponding vectors  $\mathbf{b}_i$  on both sides may be equated to give

$${}^B u_i = J_{ij} {}^A u_j \quad (6.4)$$

or in matrix form

$${}^B \mathbf{u} = \mathbf{J} {}^A \mathbf{u}. \quad (6.5)$$

This equation reduces to (6.3) when we put  $\mathbf{u} = \mathbf{a}_i$ . The matrix  $\mathbf{J}$  is a representation of the transformation from the basis  $\mathbf{A}$  to the basis  $\mathbf{B}$ ; its columns are the direction indices of  $\mathbf{a}_i$  referred to  $\mathbf{B}$ . Although this notation is often adequate, it is sometimes necessary to show explicitly that  $\mathbf{J}$  represents the operation changing the reference basis, and we then use the extended notation

$$\mathbf{J} = (\mathbf{B} \mathbf{J} \mathbf{A})$$

so that (6.5) may be written in full as

$$[\mathbf{B}; \mathbf{u}] = (\mathbf{B} \mathbf{J} \mathbf{A}) [\mathbf{A}; \mathbf{u}]. \quad (6.6)$$

The bold face symbol  $\mathbf{J}$  is to be regarded as an operator which transforms the representation  ${}^A \mathbf{u}$  into  ${}^B \mathbf{u}$ ; equivalently, we may say that  $\mathbf{J}$  is a function of  $\mathbf{A}$  and  $\mathbf{B}$  which gives the transforming matrix  $\mathbf{J}$ .<sup>†</sup> The fact that eqn. (6.6) represents a change of axes is emphasized by the use of different base symbols on either side of  $\mathbf{J}$ ; note also that the base symbols  $\mathbf{A}$  occur in juxtaposition on the right-hand side.

Provided that the vectors  $\mathbf{a}_i$ ,  $\mathbf{b}_i$  are non-coplanar sets (which is a necessary condition for them to define unit cells), the determinant  $|\mathbf{J}|$  of the matrix array  $\mathbf{J}$  does not vanish. The three simultaneous eqn. (6.5) may then be solved for  ${}^A \mathbf{u}$  in terms of  ${}^B \mathbf{u}$ , giving the reverse transformation

$${}^A \mathbf{u} = \mathbf{J}^{-1} {}^B \mathbf{u}, \quad (6.7)$$

where the *reciprocal* matrix  $\mathbf{J}^{-1}$  has elements  $J^{-1}_{ij} = J^{ji}/|\mathbf{J}|$  and  $J^{ji}$  means the cofactor of the element  $J_{ij}$  of the matrix  $\mathbf{J}$ . In the fuller matrix notation, the reverse transformation is

$$[\mathbf{A}; \mathbf{u}] = (\mathbf{A} \mathbf{J} \mathbf{B}) [\mathbf{B}; \mathbf{u}]$$

so that our notation implies

$$(\mathbf{A} \mathbf{J} \mathbf{B}) = \mathbf{J}^{-1} = (\mathbf{B} \mathbf{J} \mathbf{A})^{-1}. \quad (6.8)$$

<sup>†</sup> An alternative notation sometimes used (e.g. Bullough and Bilby, 1956) is to write  $(\mathbf{B} \mathbf{J} \mathbf{A})$  as  $(\mathbf{B}/\mathbf{A})$ . This has the slight disadvantage of not allowing the use of the single symbol  $\mathbf{J}$  when convenient.

Similarly, the transformation between sets of vectors has an inverse

$$\mathbf{B} = \mathbf{J}'^{-1} \mathbf{A}.$$

If we wish to write eqn. (6.5) in terms of row vectors, we must transpose both sides of the equation. From the rule for transposing matrix products, this gives

$${}^B \mathbf{u}' = {}^A \mathbf{u}' \mathbf{J}',$$

but if we write this in extended notation, it is convenient to have a symbol for the transpose of  $(\mathbf{B} \mathbf{J} \mathbf{A})$  which will preserve the juxtaposition of like base symbols. We thus introduce the notation

$$(\mathbf{B} \mathbf{J} \mathbf{A})' = (\mathbf{A} \mathbf{J}' \mathbf{B}) = \mathbf{J}'$$

and the transpose of (6.5) is then

$$(\mathbf{u}; \mathbf{B}) = (\mathbf{u}; \mathbf{A}) (\mathbf{A} \mathbf{J}' \mathbf{B}). \quad (6.9)$$

The new base vectors  $\mathbf{b}_i$  are associated with a new reciprocal set of vectors  $\mathbf{b}_i^*$  where

$$\mathbf{b}_i \cdot \mathbf{b}_j^* = \delta_{ij}.$$

The square of the length of  $\mathbf{u}$  may be written from (5.16) as

$$|\mathbf{u}|^2 = (\mathbf{u}; \mathbf{B}^*) [\mathbf{B}; \mathbf{u}] = (\mathbf{u}; \mathbf{B}^*) (\mathbf{B} \mathbf{J} \mathbf{A}) [\mathbf{A}; \mathbf{u}] = (\mathbf{u}; \mathbf{A}^*) [\mathbf{A}; \mathbf{u}]$$

so that

$$(\mathbf{u}; \mathbf{A}^*) = (\mathbf{u}; \mathbf{B}^*) (\mathbf{B} \mathbf{J} \mathbf{A}). \quad (6.10)$$

However,

$$(\mathbf{u}; \mathbf{A}^*) = (\mathbf{u}; \mathbf{B}^*) (\mathbf{B}^* \mathbf{J}' \mathbf{A}^*),$$

so that our notation implies

$$(\mathbf{B}^* \mathbf{J}' \mathbf{A}^*) = (\mathbf{B} \mathbf{J} \mathbf{A}) \quad \text{and} \quad (\mathbf{A} \mathbf{J}' \mathbf{B}) = (\mathbf{A}^* \mathbf{J} \mathbf{B}^*). \quad (6.11)$$

The law for the transformation of vector components referred to the reciprocal bases is thus identical with that for transforming base vectors (eqns. (6.1) and (6.2)); if  $\mathbf{J}$  is the representation of the change  $\mathbf{A} \rightarrow \mathbf{B}$ ,  $\mathbf{J}'$  is the corresponding representation of the change  $\mathbf{B}^* \rightarrow \mathbf{A}^*$ .

From eqn. (6.10), the reverse transformation may be written

$$[\mathbf{B}^*; \mathbf{u}] = (\mathbf{A} \mathbf{J}' \mathbf{B})^{-1} [\mathbf{A}^*; \mathbf{u}] = (\mathbf{B} \mathbf{J}' \mathbf{A}) [\mathbf{A}^*; \mathbf{u}]. \quad (6.12)$$

In particular, the quantities  ${}^A h_i$  must transform like this, since they are the components of  $\mathbf{h}$  referred to the basis  $\mathbf{A}^*$  (see p. 29). The transformation law is, however, usually more conveniently expressed in the form (6.10), since it is generally preferable to write the vector  $\mathbf{h}$  representing a plane normal as a row matrix  $\mathbf{h}'$ . This gives

$$(\mathbf{h}; \mathbf{B}^*) = (\mathbf{h}; \mathbf{A}^*) (\mathbf{A} \mathbf{J} \mathbf{B}) \quad (6.13)$$

or in component form

$${}^B h_j = J^{-1}_{ij} {}^A h_i.$$

We also have

$$\mathbf{h} = {}^B h_i \mathbf{b}_i^* = {}^A h_j \mathbf{a}_j^* = J_{ij} {}^B h_i \mathbf{a}_j^*,$$

so that

$$\mathbf{b}_i^* = J_{ij} \mathbf{a}_j^*$$

or

$$\mathbf{B}^* = \mathbf{J} \mathbf{A}^*, \quad (6.14)$$

which gives the relation between reciprocal vectors, corresponding to (6.2) for lattice vectors.

Finally, it is useful to examine the transformation from a base  $\mathbf{A}$  to its own reciprocal base  $\mathbf{A}^*$ . Let the components of a vector  $\mathbf{u}$  referred to the two bases be related by

$$\mathbf{u}^* = \mathbf{G} \mathbf{u}, \quad (6.15)$$

where  $\mathbf{G} = (\mathbf{A}^* \mathbf{G} \mathbf{A})$  is the matrix representation of the transformation from  $\mathbf{A}$  to  $\mathbf{A}^*$ . From eqn. (5.15) the scalar product of two vectors may be written

$$\mathbf{u} \cdot \mathbf{v} = u_j^* v_j = G_{ji} u_i v_j,$$

but this product may also be expanded as

$$\mathbf{u} \cdot \mathbf{v} = u_i \mathbf{a}_i \cdot v_j \mathbf{a}_j = u_i v_j \mathbf{a}_i \cdot \mathbf{a}_j$$

so that the elements of the matrix  $\mathbf{G}$  are

$$G_{ij} = G_{ji} = \mathbf{a}_i \cdot \mathbf{a}_j. \quad (6.16)$$

The function symbol  $\mathbf{G}$  is called the metric,<sup>†</sup> and is represented by the symmetrical square matrix  $\mathbf{G}$ . Its importance arises from its fundamental connection with the distance between two points, since the square of the length of a vector is

$$|\mathbf{u}|^2 = \mathbf{u}' \mathbf{u}^* = \mathbf{u}' \mathbf{G} \mathbf{u} = G_{ij} u_i u_j. \quad (6.17)$$

The reverse transformation is

$$\mathbf{u} = \mathbf{G}^{-1} \mathbf{u}^*, \quad (6.18)$$

where  $\mathbf{G}^{-1}$  has components  $G^{-1}_{ij} = G^{-1}_{ji} = \mathbf{a}_i^* \cdot \mathbf{a}_j^*$ . The length of  $\mathbf{u}$  may then also be expressed as

$$|\mathbf{u}|^2 = \mathbf{u}^{*'} \mathbf{G}^{-1} \mathbf{u}^* = G^{-1}_{ij} u_i^* u_j^*.$$

Similarly, eqn. (5.22) gives for the interplanar spacing  $d$  of the planes  $\mathbf{h}$

$$(d)^{-2} = |\mathbf{h}|^2 = G^{-1}_{ij} h_i h_j. \quad (6.19)$$

<sup>†</sup> Since the relation between  $\mathbf{A}$  and  $\mathbf{A}^*$  is defined by  $\mathbf{A}$  alone,  $\mathbf{G}$  may also be regarded as a second order tensor (see next section) and the matrix  $\mathbf{G} = G_{ij}$  then gives the components of this tensor in the system  $\mathbf{A}$ .  $\mathbf{G}$  is usually called the metric tensor. The function symbol  $\mathbf{J}$ , connecting arbitrary axis systems, is not a tensor.

It follows from a result to be proved later (Section 7, p.47-8) that the determinant of the matrix  $\mathbf{G}$  is equal to the ratio of the volume of the cell defined by  $\mathbf{a}_i$  to that defined by  $\mathbf{a}_i^*$ . Combining this result with eqn. (5.12) shows that

$$|\mathbf{G}|^{1/2} = v_a \quad \text{and similarly} \quad |\mathbf{G}^{-1}|^{1/2} = v_a^*, \quad (6.20)$$

where, as before,  $v_a, v_a^*$  are the volumes of the respective unit cells.

Since  $\mathbf{G}$  is symmetric,  $\mathbf{G}' = \mathbf{G}$ , or in full

$$(\mathbf{A} \mathbf{G}' \mathbf{A}^*) = (\mathbf{A}^* \mathbf{G} \mathbf{A}).$$

We now write the scalar product of  $\mathbf{u}$  and  $\mathbf{v}$  in the basis  $\mathbf{A}$  and transform to the basis  $\mathbf{B}$  using eqns. (6.17) and (6.6)

$$\mathbf{u} \cdot \mathbf{v} = (\mathbf{u}; \mathbf{A}) (\mathbf{A}^* \mathbf{G} \mathbf{A}) [\mathbf{A}; \mathbf{v}] = (\mathbf{u}; \mathbf{B}) (\mathbf{B} \mathbf{J}' \mathbf{A}) (\mathbf{A}^* \mathbf{G} \mathbf{A}) (\mathbf{A} \mathbf{J} \mathbf{B}) [\mathbf{B}; \mathbf{v}]$$

so that

$$(\mathbf{B}^* \mathbf{G} \mathbf{B}) = (\mathbf{B} \mathbf{J}' \mathbf{A}) (\mathbf{A}^* \mathbf{G} \mathbf{A}) (\mathbf{A} \mathbf{J} \mathbf{B}), \quad (6.21)$$

which gives the relation between the representations of  $\mathbf{G}$  in two different bases. We could also have derived (6.21) by making use of the identity  $(\mathbf{B}^* \mathbf{G} \mathbf{B}) = (\mathbf{B}^* \mathbf{J} \mathbf{A}^*) (\mathbf{A}^* \mathbf{G} \mathbf{A}) (\mathbf{A} \mathbf{J} \mathbf{B})$  and substituting for  $(\mathbf{B}^* \mathbf{J} \mathbf{A}^*)$  from (6.11).

Consider now the orthonormal basis  $\mathbf{i}$ . The basic vectors  $\mathbf{i}_i$  then satisfy eqn. (5.2) and  $\mathbf{G}^{-1} = \mathbf{G} = \mathbf{I}$ , the unit matrix having elements  $I_{ij} = \delta_{ij}$ . The metric of an orthonormal system is thus unity. This corresponds to the fact that the bases  $\mathbf{i}$  and  $\mathbf{i}^*$  are identical in such a system.

Examination of the above equations shows that the laws of transformation are of two kinds. Quantities which transform like the vector components  $u_i$  or the reciprocal base vectors  $\mathbf{a}_i^*$  are called *contravariant* in tensor analysis; those which transform like the base vectors  $\mathbf{a}_i$  or the plane indices (reciprocal vector components)  $h_i$  are called *covariant*. In addition, there are scalar quantities which are *invariant* with respect to an axis transformation. Thus  $u_i$  form the components of a contravariant tensor of the first order (a vector), or more simply, may be described as the contravariant components of the real vector  $\mathbf{u}$ ;  $u_i^*$  are the covariant components of  $\mathbf{u}$ . Now let  ${}^A u_i, {}^B u_i$  represent continuous variables along the axes  $\mathbf{a}_i, \mathbf{b}_i$ , so that they define coordinate systems. The linear relation between the coordinates, given by (6.4), may be written

$${}^B u_i = (\partial {}^B u_i / \partial {}^A u_j) {}^A u_j, \quad (6.22)$$

where  $J_{ij} = \partial {}^B u_i / \partial {}^A u_j$ , etc. This is the law for the transformation of the contravariant components of  $\mathbf{u}$ . The corresponding law for the covariant components  $u_i^*$  is

$${}^B u_i^* = (\partial {}^A u_j / \partial {}^B u_i) {}^A u_j^*. \quad (6.23)$$

In tensor notation, contravariant quantities are distinguished by writing the identifying suffix as a superscript, and there is then no need for a separate notation for the bases  $\mathbf{A}$  and  $\mathbf{A}^*$ . Thus eqn. (5.1) would be written

$$\mathbf{u} = u^i \mathbf{a}_i, \quad (6.24)$$

where the summation convention applies as before. Note that the superscript  $i$  is not a power index.

Covariant quantities are written with the suffix as subscript, so that the covariant components  $u_i^*$  are now written simply  $u_i$ . From eqn. (5.13) it follows that the covariant components of  $\mathbf{u}$  in base  $A$  are equal to the orthogonal projections of  $\mathbf{u}$  on the axes  $\mathbf{a}_i$  (see Fig. 2.1). From (5.12) we also have

$$\mathbf{u} = u_i \mathbf{a}^i, \quad (6.25)$$

where  $\mathbf{a}^i$  are the reciprocal vectors, formerly written  $\mathbf{a}_i^*$ . The covariant components of  $\mathbf{u}$  in the base  $A$  are thus the contravariant components in the base  $A^*$ , and vice versa. In general, we may alter any equation of the form (6.24) by lowering the dummy index in one place and raising it in another.

In terms of the new notation, the transformation laws (6.22) and (6.23) become

$${}^B u^i = (\partial^B u^i / \partial^A u^j) {}^A u^j \quad (6.26)$$

$$\text{and} \quad {}^B u_i = (\partial^A u^j / \partial^B u^i) {}^A u_j = (\partial^B u_i / \partial^A u_j) {}^A u_j. \quad (6.27)$$

In the above discussion, we have been careful to write about the covariant and contravariant components of one vector  $\mathbf{u}$ , since we wish to emphasize the idea of the vector as the physical entity. In tensor algebra, it is customary to treat the components as separate covariant and contravariant vectors, and this leads to economy of description. The vector  $\mathbf{u}$  is sometimes called the real vector.

Anticipating the results of the next section a little, we may also form covariant, contravariant and mixed tensors of the second order, with components  $T_{ij}$ ,  $T^{ij}$ , and  $T_j^i$  respectively. These tensors are arrays of nine quantities, each depending on two directions of the coordinate axes, which represent a linear relation between two vectors (e.g. the stress tensor relates the linear force on a surface element and the vector normal to the element). When the vectors are referred to another coordinate system, the quantities in the tensor transform according to laws which are contravariant for both suffices, covariant for both suffices, or mixed, thus:

$$\left. \begin{array}{l} \text{(Contravariant)} \\ \text{(Covariant)} \\ \text{(Mixed)} \end{array} \right\} \begin{array}{l} {}^B T^{ij} = (\partial^B u^i / \partial^A u^k) (\partial^B u^j / \partial^A u^l) {}^A T^{kl}, \\ {}^B T_{ij} = (\partial^A u^k / \partial^B u^i) (\partial^A u^l / \partial^B u^j) {}^A T_{kl} \\ \quad = (\partial^B u_i / \partial^A u_k) (\partial^B u_j / \partial^A u_l) {}^A T_{kl}, \\ {}^B T_j^i = (\partial^B u^i / \partial^A u^k) (\partial^A u^l / \partial^B u^j) {}^A T_l^k \\ \quad = (\partial^B u^i / \partial^A u^k) (\partial^B u_j / \partial^A u_l) {}^A T_l^k. \end{array} \quad (6.28)$$

In the remainder of this book we shall not use tensor notation, preferring the matrix representation when it is necessary to distinguish covariance from contravariance. When components are required, they will therefore be written with all suffices as subscripts, except for the standard notation for the cofactor of a matrix element, used previously. Tensor notation is especially powerful in problems where curvilinear coordinates are required, so that the linear relations (6.2) and (6.4) have to be replaced by the more general

$${}^B u^i = f^i({}^A u^1, {}^A u^2, {}^A u^3)$$

so that

$$d^B u^i = (\partial^B u^i / \partial^A u^j) d^A u^j. \quad (6.29)$$

The general properties of contravariance or covariance are then still defined by eqns. (6.26)–(6.28), but it should be noted that the coordinates  $u^i$  themselves no longer transform according to the contravariant law. The transformation of coordinates is only contravariant when the relation (6.29) is linear of type (6.4); this is called an affine transformation (see next section) and the coordinate systems are all Cartesian. In all other cases, it is clearly to some extent arbitrary whether we write the coordinates as  $u^i$  or  $u_i$ .

It is perhaps of interest to show that the components of the matrix  $\mathbf{G}$  may be regarded as the representation of a tensor. If the coordinates of two neighbouring points in the reference system A are  $^A u^i$  and  $^A u^i + d^A u^i$ , we know that the separation of the points,  $dw$ , is given by

$$(dw)^2 = {}^A G_{ij} d^A u^i d^A u^j.$$

In the system B, the corresponding equation is given from (6.29) as

$$(dw)^2 = {}^B G_{ij} d^B u^j d^B u^i = {}^B G_{ij} (\partial^B u^i / \partial^A u^k) (\partial^B u^j / \partial^A u^l) d^A u^k d^A u^l,$$

and since the separation is an invariant,

$${}^A G_{ij} = (\partial^B u^i / \partial^A u^k) (\partial^B u^j / \partial^A u^l) {}^B G_{kl}. \quad (6.30)$$

Comparison with (6.28) shows that the components of  $\mathbf{G}$  transform according to the law for covariant tensors of the second order, so that we are justified in regarding  $\mathbf{G}$  as a tensor. In effect,  $\mathbf{G}$  relates two vectors such that the contravariant components of the second are the covariant components of the first. In the same way, we can show that the components of the matrix  $\mathbf{G}^{-1}$  transform according to the law for contravariant tensors (6.28), and would be written  $G^{ij}$  in tensor notation.

Note that if both sets of axes are orthogonal and Cartesian, the covariant and contravariant components coincide, and the two transformation laws are identical. This is why contravariance and covariance are not distinguished in elementary vector analysis. We shall frequently use oblique Cartesian coordinates, but curvilinear coordinates are not needed for most problems in crystal geometry. As already emphasized, we shall now employ matrix rather than tensor notation.

For reference, we give in Table II a summary comparison of the main features of the different notations, and of the standard equations of crystal geometry expressed in these notations. Most of the results in the table have been obtained in the text above, but there is one additional point to note.

Equations (6.17) and (6.19) for the length of a vector and the spacing of a set of planes respectively are valid for any choice of the base vectors  $\mathbf{a}_i$ , including vectors which do not define a primitive unit cell. However, we may need to know the distance between adjacent lattice points (identity distance) in the direction  $\mathbf{u}$ , or between adjacent lattice planes normal to  $\mathbf{h}$ . If we use components  $u_i, h_i$  which are integers with no common factor, these dis-

TABLE II. SUMMARY OF LATTICE GEOMETRY

	Matrix notation	Tensor notation
Base vectors	$\mathbf{A} = \mathbf{a}_i$	$\mathbf{a}_i$
Reciprocal base vectors	$\mathbf{A}^* = \mathbf{a}_i^*$	$\mathbf{a}^i$
Contravariant components of $\mathbf{u}$	$\mathbf{u} = [\mathbf{A}; \mathbf{u}]$	$u^i$
Covariant components of $\mathbf{u}$	$\mathbf{u}^* = [\mathbf{A}^*; \mathbf{u}]$	$u_i$
Covariant components of plane normal, $\mathbf{h}$	$h'$ (strictly $h^{**}$ ) $= (\mathbf{h}; \mathbf{A}^*)$	$h_i$
Metric $\mathbf{a}_i \cdot \mathbf{a}_j$	$\mathbf{G} = (\mathbf{A}^* \mathbf{G} \mathbf{A})$	$G_{ij}$
Reciprocal metric, $\mathbf{a}^i \cdot \mathbf{a}^j$	$\mathbf{G}^{-1} = (\mathbf{A} \mathbf{G}^{-1} \mathbf{A}^*)$	$G^{ij}$
Volume of cell, $v_a$	$ \mathbf{G} ^{1/2}$	
Volume of reciprocal cell, $v_a^*$	$ \mathbf{G}^{-1} ^{1/2}$	
Scalar product, $\mathbf{u} \cdot \mathbf{v}$	$\mathbf{u}' \mathbf{v}^* = \mathbf{u}' \mathbf{G} \mathbf{v}$ $= \mathbf{u}^{**} \mathbf{G}^{-1} \mathbf{v}^*$	$G_{ij} u^i v^j = G^{ij} u_i v_j$
Contravariant components of $\mathbf{w} = \mathbf{u} \wedge \mathbf{v}$		$w^i =  \mathbf{G}^{-1} ^{1/2} \varepsilon^{ijk} u_j v_k \dagger$
Covariant components of $\mathbf{w} = \mathbf{u} \wedge \mathbf{v}$		$w_i =  \mathbf{G} ^{1/2} \varepsilon_{ijk} u^j v^k \dagger$
Repeat distance along $\mathbf{u}$ ( $u^i$ relatively prime)	$I'(\mathbf{u}' \mathbf{G} \mathbf{u})^{1/2}$	$I'(G_{ij} u^i u^j)^{1/2}$
Interplanar spacing, $d$ of planes with $h_i$ relatively prime	$I(h' \mathbf{G}^{-1} \mathbf{h})^{-1/2}$	$I(G^{ij} h_i h_j)^{-1/2}$
Cosine of angle between $\mathbf{u}$ and $\mathbf{v}$	$\frac{\mathbf{u}' \mathbf{G} \mathbf{v}}{(\mathbf{u}' \mathbf{G} \mathbf{u})^{1/2} (\mathbf{v}' \mathbf{G} \mathbf{v})^{1/2}}$	$\frac{G_{ij} u^i v^j}{(G_{ij} u^i u^j)^{1/2} (G_{ij} v^i v^j)^{1/2}}$
Cosine of angle between planes $\mathbf{h}$ and $\mathbf{k}$	$\frac{h' \mathbf{G}^{-1} \mathbf{k}}{(h' \mathbf{G}^{-1} \mathbf{h})^{1/2} (k' \mathbf{G}^{-1} \mathbf{k})^{1/2}}$	$\frac{G^{ij} h_i k_j}{(G^{ij} h_i h_j)^{1/2} (G^{ij} k_i k_j)^{1/2}}$
Cosine of angle between direction $\mathbf{u}$ and plane $\mathbf{h}$	$\frac{h' \mathbf{u}}{(h' \mathbf{G}^{-1} \mathbf{h})^{1/2} (\mathbf{u}' \mathbf{G} \mathbf{u})^{1/2}}$	$\frac{u^i h_j}{(G_{ij} u^i u^j)^{1/2} (G^{ij} h_i h_j)^{1/2}}$
Zone axis of $\mathbf{h}$ and $\mathbf{k}$	$\mathbf{u}$ parallel $\mathbf{h} \wedge \mathbf{k}$	$u^i \propto$ to $\varepsilon^{ijk} h_j h_k$
Plane containing $\mathbf{u}$ and $\mathbf{v}$	$\mathbf{h}$ parallel $\mathbf{u} \wedge \mathbf{v}$	$h_i \propto$ to $\varepsilon_{ijk} u^j v^k$

† The factors  $|\mathbf{G}|^{1/2}$  and  $|\mathbf{G}^{-1}|^{1/2}$  are often included in the definition of  $\varepsilon_{ijk}$  and  $\varepsilon^{ijk}$  respectively.

tances are given by the same expressions, (6.17) and (6.19), provided the basis  $\mathbf{A}$  is primitive. As we have already stated, however, it is often convenient to use centred (non-primitive) cells, so that the vectors  $u_i \mathbf{a}_i$  with  $u_i$  taking all possible integral values do not represent *all* the lattice points. For example, in the body-centred cubic structure with the conventional cubic basis, the lattice points are given by the vectors  $\frac{1}{2} u_i \mathbf{a}_i$ , with the restriction that the quantities  $u_i$  must be all odd or all even integers. This complication is allowed for in Table II by introducing the "cell factors"  $I, I'$  into the equations for identity distance and planar spacing. The cell factors for the important structures are given separately in Table III.

TABLE III. CELL FACTORS FOR CENTRED LATTICES

Type of cell	Plane indices	$I$	Direction indices	$I'$
Primitive	All $h_i$	1	All $u_i$	1
Body-centred	$\Sigma h_i$ even	1	$u_i$ all odd	$\frac{1}{2}$
	$\Sigma h_i$ odd	$\frac{1}{2}$	$u_i$ mixed odd and even	1
Base-centred	$h_1 + h_2$ even	1	$u_1 + u_2$ even	$\frac{1}{2}$
	$h_1 + h_2$ odd	$\frac{1}{2}$	$u_1 + u_2$ odd	1
Face-centred	$h_i$ all odd	1	$\Sigma u_i$ even	$\frac{1}{2}$
	$h_i$ mixed odd and even	$\frac{1}{2}$	$\Sigma u_i$ odd	1

We conclude this section by referring to the well-known Miller–Bravais four-axis system for indexing directions and planes in hexagonal structures. The four-axis convention is described in most elementary textbooks on crystallography; it has the advantage that symmetry-related planes and directions are obtained by a simple permutation of indices, but it is clearly inconvenient when a hexagonal lattice has to be related to a lattice of different symmetry. However, the relations between the direct and reciprocal lattices when a four-axis system is used, together with associated problems of crystal geometry, are not obvious, and we shall briefly describe some aspects of these relations in order to complete our description of lattice geometry.

The hexagonal lattice may be regarded as a series of planar hexagonal nets of edge  $a$  stacked vertically above each other with a separation  $c$ . The most useful bases of conventional type are defined by the vector sets  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{c}$  and  $\mathbf{a}_1, \mathbf{a}_1 + 2\mathbf{a}_2, \mathbf{c}$ , which describe respectively a primitive cell and a C-centred orthorhombic cell. Various rhombohedral bases may also be used, but for hexagonal lattices their disadvantages outweigh their advantages. The non-conventional four-axis basis consists of the vectors  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{c}$  and a vector direction  $\mathbf{u}$  is represented by

$$\mathbf{u} = u_i \mathbf{a}_i \quad (i = 1, \dots, 4), \quad (6.31)$$

where it is a requirement that

$$u_1 + u_2 + u_3 = 0. \quad (6.32)$$

Rational lattice directions are represented by relatively prime integral values of  $u_i$ , and the coordinates of all the lattice points along such a direction through the origin are then given by  $I'nu_i$ , where  $n$  is any integer and  $I'$  is equal to  $\frac{1}{3}$  when  $u_1 - u_2$  is divisible by 3 and otherwise is unity. ( $I'$  is analogous to the cell factors of Table II.)

If lattice planes are now defined in terms of their reciprocal intercepts  $h_i$  on the axes of the direct basis, we find correspondingly that

$$h_1 + h_2 + h_3 = 0. \quad (6.33)$$

However, some complications arise when we represent plane normals as directions of the reciprocal lattice. With a four-axis system it is not possible to define a reciprocal basis by

means of eqns. (5.11), since no vector can be simultaneously perpendicular to three non-coplanar vectors. Although a suitable four-axis basis for the reciprocal lattice can readily be found, the fact that this basis is not strictly reciprocal to the direct basis has caused much confusion in the literature.

Clearly the reciprocal lattice itself is completely defined by the direct lattice; it may be introduced through eqns. (5.11) applied to a conventional direct basis, or by using the geometrical interpretation discussed at the end of Section 5, which is independent of any choice of basis. The reciprocal lattice is also hexagonal and consists of planar nets of edge  $(2/3)^{-2}a^{-1}$  stacked at a separation of  $c^{-1}$ . The hexagonal nets of the direct and reciprocal lattices have parallel normals, but are rotated through  $30^\circ$  relative to each other. The appropriate four-axis reciprocal basis is defined by the vectors

$$\mathbf{a}_i^+ = \left(\frac{2}{3}a^{-2}\right)\mathbf{a}_i \quad (i = 1, \dots, 3), \quad \mathbf{a}_4^+ = c^{-2}\mathbf{a}_4, \quad (6.34)$$

where, following Nicholas and Segall (1970), we have used the notation  $\mathbf{a}_i^+$  rather than  $\mathbf{a}_i^*$  because eqns. (5.11) are not valid. The normal to the plane  $h_i$  is now a vector  $\mathbf{h}$  of the reciprocal lattice, where

$$\mathbf{h} = h_i\mathbf{a}_i^+ \quad (i = 1, \dots, 4). \quad (6.35)$$

This equation is strictly analogous to (5.19). It also follows that if the direction  $u_i$  is contained in the plane  $h_i$ , then

$$h_i u_i = 0 \quad (i = 1, \dots, 4), \quad (6.36)$$

which is the obvious analogue of the three-axis condition.

It should be noted that the four-axis basis of the direct lattice defines a hexagonal prism, of volume  $3^{\frac{3}{2}} a^2 c / 2$ , which contains three lattice points; this is why the cell factor  $I'$  has to be introduced when repeat distances are calculated. However, the vectors  $\mathbf{a}_i^+$  define a hexagonal prism of the reciprocal lattice which contains only one reciprocal lattice point, and the reciprocal spacings of lattice planes are therefore given by  $|\mathbf{h}|$ , where  $h_i$  are relatively prime integers, without the need to introduce a cell factor  $I$ .

Since eqns. (5.11) are invalid, the existence of the reciprocal basis (6.34) for which eqns. (6.35) and (6.36) are satisfied, appears almost fortuitous. That this is not so was shown by Frank (1965) and Nicholas and Segall (1970). Frank's approach is to consider a four-dimensional lattice with an orthogonal basis  $\alpha_i$  with  $|\alpha_i| = \alpha$  for  $i = 1, \dots, 3$  and  $|\alpha_4| = \gamma$ . Equations (5.11), with a range of  $1, \dots, 4$ , for  $i$ , then enable an orthogonal reciprocal basis  $\alpha_i^*$  to be defined. The real three-dimensional lattice is obtained either by projecting the four-dimensional lattice along some direction or by sectioning it in a particular "hyperplane". Frank showed that a projection along [1110] gives the hexagonal lattice with parameters  $a = (\frac{2}{3})^{1/2} \alpha$  and  $c = \gamma$ . Similarly, a section of the four-dimensional reciprocal lattice in (1110) gives the reciprocal of the real hexagonal lattice. Moreover the four-axis bases of the direct and reciprocal lattices are simply obtained by projecting the orthogonal bases  $\alpha_i$  and  $\alpha_i^*$ . The significance of the above procedure is readily appreciated by lowering the dimensionality by one; it then becomes equivalent to projecting a simple cubic lattice along [111] or sectioning its reciprocal lattice in (111) to give the direct and reciprocal nets, and

the three-axis bases in these nets are obtained from the projections of the cubic direct and reciprocal axes.

Nicholas and Segall (1970) have given a more complete description of the general case of a redundant base vector, i.e. the use in an  $n$ -dimensional space of a basis  $\mathbf{a}_i$  with  $i = 1, \dots, n+1$ . They show that (6.36) is universally valid (whatever  $i$ ) provided that  $h_i$  are defined by the reciprocal intercepts of a hyperplane on  $\mathbf{a}_i$ , and also that it is always possible to define a reciprocal basis  $\mathbf{a}_i^+$  for which (6.35) is valid. Moreover, this choice is not unique; in the hexagonal lattice, for example, any other basis  $\mathbf{a}_i^{\#} = \mathbf{a}_i^+ + \mathbf{d}^+$  ( $i = 1, \dots, 3$ ),  $\mathbf{a}_4^{\#} = \mathbf{a}_4^+$ , where  $\mathbf{d}^+$  is any vector, will not disturb the validity of (6.35). The real justification for the choice of axes which is customarily made (and hence for the conditions (6.32) and (6.33)) is in order to ensure that cyclically permuted indices represent crystallographically equivalent directions and planes.

The advantage of linking the four-axis systems to the orthogonal vectors of a four-dimensional lattice is that crystallographic formulae can be derived rather simply for the four-dimensional lattice and then transformed to the four-axis representation of a three-

TABLE IIIA. SUMMARY OF LATTICE GEOMETRY WITH MILLER-BRAVAIS INDICES

Base vectors	$\mathbf{a}_i \quad (i = 1, \dots, 4)$
Reciprocal lattice basis	$\mathbf{a}_i^+ \quad (i = 1, \dots, 4)$
Direction vector, $\mathbf{u}$	$u_i \mathbf{a}_i$
Plane normal, $\mathbf{h}$	$h_i \mathbf{a}_i^+$
Metric of four-space, $\mathbf{G}$	$G_{ij}^{-1} = \text{diag} \left\{ \frac{3}{2}a^2, \frac{3}{2}a^2, \frac{3}{2}a^2, c^2 \right\}$
Reciprocal metric, $\mathbf{G}^{-1}$	$G_{ij} = \text{diag} \left\{ \frac{2}{3}a^{-2}, \frac{2}{3}a^{-2}, \frac{2}{3}a^{-2}, c^{-2} \right\}$
Volume of hexagonal prism (3 lattice points) defined by $\mathbf{a}_i$	$3^{3/2} a^2 c / 2$
Volume of primitive hexagonal prism defined by $\mathbf{a}_i^+$	$\left(\frac{2}{3}\right)^{1/2} a^{-2} c^{-1}$
Scalar product, $\mathbf{u} \cdot \mathbf{v}$	$G_{ij} u_i v_j$
Repeat distance along $\mathbf{u}$ ( $u_i$ relatively prime)	$I'(G_{ij} u_i u_j)^{1/2}$
"Cell factor"	$I' = \frac{1}{3}$ if $u_1 - u_2 = 3n$ $I' = 1$ if $u_1 - u_2 \neq 3n$
Interplanar spacing ( $h_i$ relatively prime)	$(G_{ij}^{-1} h_i h_j)^{-1/2}$
Cosine of angles between $\mathbf{u}$ and $\mathbf{v}$ , $\mathbf{h}$ and $\mathbf{k}$ , $\mathbf{u}$ and $\mathbf{h}$	See Table II - the equations are identical with $i, j = 1, \dots, 4$
Zone axis, $\mathbf{u}$ , of $\mathbf{h}$ and $\mathbf{k}$	$u_1 \propto h_4(k_2 - k_3) - k_4(h_2 - h_3)$ $u_2 \propto h_4(k_3 - k_1) - k_4(h_3 - h_1)$ $u_3 \propto h_4(k_1 - k_2) - k_4(h_1 - h_2)$ $u_4 \propto -3(h_1 k_2 - k_1 h_2)$
Normal, $\mathbf{h}$ , to plane containing $\mathbf{u}$ and $\mathbf{v}$	$h_1 \propto u_4(v_2 - v_3) - v_4(u_2 - u_3)$ $h_2 \propto u_4(v_3 - v_1) - v_4(u_3 - u_1)$ $h_3 \propto u_4(v_1 - v_2) - v_4(u_1 - u_2)$ $h_4 \propto -3(u_1 v_2 - v_1 u_2)$

dimensional lattice. Particular equations (not usually derived in this way) are given in various papers (e.g. Otte and Crocker, 1965, 1966; Nicholas, 1966; Neumann, 1966; Okamoto and Thomas, 1968); most of these papers contain mistakes as pointed out by Nicholas (1970). The most important formulae are summarized in Table IIIA, which may be looked upon as a supplement to Table II. By introducing the metric of the four-dimensional space, formulae containing scalar products may be written in vector form, exactly as in Table II. Vector products, however, cannot be expressed quite so simply because the two bases are not properly reciprocal, and we have therefore written out the components of two typical vector products in full. The full tensor notation with contravariance and covariance distinguished by superscript and subscript indices has not been used in Table IIIA, but may readily be derived as in Table II.

Finally, it should perhaps be emphasized that the whole of this descriptions is concerned with the geometry of the hexagonal *lattice*. The hexagonal close-packed (h.c.p.) structure is derived from that lattice by placing two atoms around each lattice point, e.g. in sites  $0, 0, 0, 0$  and  $\frac{1}{3}, 0, -\frac{1}{3}, \frac{1}{2}$  and their equivalents (see footnote on p. 122).

#### 7. AFFINE TRANSFORMATIONS: HOMOGENEOUS DEFORMATION

Equations (6.5) or (6.7) are commonly interpreted in two ways. In the first of these, used in the last section, the quantities  ${}^A u_i, {}^B u_i$  are the components of the same vector  $\mathbf{u}$  referred to two different sets of base vectors. An alternative interpretation is to suppose that we have a fixed reference system  $\mathbf{a}_i$ , and the equations then represent a physical transformation which changes a vector  ${}^A \mathbf{u}$  into another vector  ${}^B \mathbf{u}$ , where  ${}^B u_i$  are the components of  ${}^B \mathbf{u}$  in the fixed reference system. This second interpretation may be used, for example, to specify the relations between two different crystal lattices, which are in fixed orientations with respect to each other. The possibility of interpreting (6.5) in these two ways is a result of the linear nature of both axis transformations and homogeneous deformations. Nevertheless, some confusion may arise if the equation is freely interpreted in either sense, as is sometimes done, and one advantage of the extended matrix notation is that it completely avoids ambiguity of this kind.

Consider the general linear transformation

$$\mathbf{v} = \mathbf{S} \mathbf{u}, \quad (7.1)$$

where  $\mathbf{S}$  is a physical entity (a tensor) which converts the vector  $\mathbf{u}$  into a new vector  $\mathbf{v}$ . In matrix notation, this equation is written

$$\mathbf{v} = \mathbf{S} \mathbf{u} \quad \text{or} \quad [\mathbf{A}; \mathbf{v}] = (\mathbf{A} \mathbf{S} \mathbf{A}) [\mathbf{A}; \mathbf{u}]. \quad (7.2)$$

The extended notation emphasizes that all physical quantities,  $\mathbf{u}$ ,  $\mathbf{v}$ , and  $\mathbf{S}$  are referred to the basis  $\mathbf{A}$ , and a clear distinction is obtained between square matrix quantities of the type  $(\mathbf{A} \mathbf{S} \mathbf{A})$ , which are the representations of a tensor in some particular coordinate system, and those of type  $(\mathbf{B} \mathbf{S} \mathbf{A})$ , which are the representations of a function operator connecting two coordinate systems.

During a transformation of type (7.1), points which were originally collinear remain collinear, and lines which were originally coplanar remain coplanar. Such a transformation is called affine; it represents a homogeneous deformation of space, or of the crystal lattice.

Equation (7.2) is the matrix representation of a homogeneous deformation referred to the system  $A$ . In the basis  $B$ , there will be a corresponding representation

$$[B; v] = (B S B) [B; u]. \quad (7.3)$$

Suppose the relation between  $A$  and  $B$  is given by

$$[B; u] = (B J A) [A; u]$$

so that (7.2) may be written

$$(A J B) [B; v] = (A S A) (A J B) [B; u]$$

and multiplying both sides by  $(B J A)$

$$[B; v] = (B J A) (A S A) (A J B) [B; u].$$

Comparing this equation with (7.3) we see that

$$(B S B) = (B J A) (A S A) (A J B), \quad (7.4)$$

and this is called the similarity transform of  $(A S A)$  into  $(B S B)$ , both of these matrices being representations of the tensor  $S$ . In shortened form, the equation is

$${}^B S = J {}^A S J^{-1}.$$

The usefulness of the juxtaposition of the bases in the extended form of the equation should be noted.<sup>†</sup>

Now consider  $u$  to be any vector in the plane having normal  $h$  so that  $h \cdot u = h' u = 0$ . After deformation, the vector  $u$  is changed into a vector  $v$ , and the plane has a new normal  $k$  where  $k \cdot v = 0$ . Writing the matrix representations of these two equations in the basis  $A$ , we have

$$h' u = k' S u = 0 \quad \text{or} \quad k' = h' S^{-1}. \quad (7.5)$$

This gives the effect of the tensor  $S$  on the components of vectors normal to planes, and may be written in full as

$$(k; A^*) = (h; A^*) (A S A)^{-1}.$$

Let us now consider the properties of the most general form of homogeneous deformation. The following statements will be taken as self-evident, though formal proofs occur incidentally later in this section. If we imagine a sphere inscribed in the material before deformation, this would be distorted into a triaxial ellipsoid, called the strain ellipsoid. The principal axes of the strain ellipsoid would be mutually perpendicular before deformation, and would thus have suffered no relative change in orientation, although in general

<sup>†</sup> It will be noted that the components of the matrix  $S$  are those of a mixed tensor, and the tensor form of (7.2) is  $v' = S^j u^i$ .

each would have been rotated from its original position, and changed in length. It follows also that we could have inscribed an ellipsoid in the material before deformation, such that after deformation it became a sphere; this is called the reciprocal strain ellipsoid. Lines in the directions of the axes of the reciprocal strain ellipsoid before deformation lie in the directions of the axes of the strain ellipsoid after deformation; the axes of the reciprocal strain ellipsoid are called the directions of principal strain.

In the most general deformation, all vectors change their length, but there is at least one vector (and generally three) which is unrotated. We may prove this as follows. Suppose for some vector  $\mathbf{u}$  the transformation leaves  $\mathbf{v}$  parallel to  $\mathbf{u}$ . Then the only effect of  $\mathbf{S}$  is to multiply the components of  $\mathbf{u}$  by a constant scalar factor, say  $\lambda_i$ . Then

$$\mathbf{S} \mathbf{u} = \lambda_i \mathbf{u} \quad \text{or} \quad (\mathbf{S} - \lambda_i \mathbf{I}) \mathbf{u} = 0, \quad (7.6)$$

where  $\mathbf{I}$  is the unit matrix. This equation has non-trivial solutions ( $u_i \neq 0$  for all  $i$ ) only when

$$|\mathbf{S} - \lambda_i \mathbf{I}| = 0. \quad (7.7)$$

Equation (7.7) is a cubic in  $\lambda_i$ , and is called the characteristic equation of the matrix  $\mathbf{S}$ . If  $\mathbf{S}$  is real, there are three roots, of which one must necessarily be real; if the matrix  $\mathbf{S}$  is symmetric, all three must be real. There is thus always one possible solution of (7.6), giving a vector which is unchanged in direction (if  $\lambda_i$  is negative, it is reversed in sign, but this does not correspond to a physically achievable deformation). If all three roots are real, there are three such directions. The values of the roots, which are called the eigenvalues of the matrix, are given by the following equations, obtained by expanding (7.7):

$$\left. \begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= S_{11} + S_{22} + S_{33} = S_{ii}, \\ \lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_3 \lambda_1 &= S_{11} S_{22} + S_{22} S_{33} + S_{33} S_{11} - S_{12} S_{21} - S_{23} S_{32} - S_{31} S_{13} \\ \lambda_1 \lambda_2 \lambda_3 &= |\mathbf{S}|. \end{aligned} \right\} \quad (7.8)$$

It is readily proved that two matrices related by a similarity transformation have the same eigenvalues. Since the eigenvalues of a diagonal matrix are simply its non-zero components, it follows that provided the matrix has three distinct, real eigenvalues, it can always be reduced to diagonal form by a similarity transformation. Physically, this corresponds to an axis transformation to a new set of coordinates lying along the unrotated directions. It is obvious that the matrix representation of  $\mathbf{S}$  in this system will be a diagonal matrix.

Having found the eigenvalues, we can determine an axis transformation which will diagonalize  $\mathbf{S}$  as follows. Choose the first root,  $\lambda_1$ , and write (7.6) in the form:

$$(\mathbf{S} - \lambda_1 \mathbf{I}) \bar{\mathbf{a}}_1 = 0 \quad \text{or} \quad \mathbf{E} \bar{\mathbf{a}}_1 = 0, \quad (7.9)$$

where  $\bar{\mathbf{a}}_1$  is the matrix representation of one of the unrotated vectors, which we now call  $\bar{\mathbf{a}}_1$ . Since (7.9) represents three linear homogeneous equations, we cannot determine the components  $(\bar{a}_1)_i$  of  $\bar{\mathbf{a}}_1$  uniquely, but we may find their ratio. If eqns. (7.9) are written in full, we see that a possible solution is

$$(\bar{a}_1)_i = C E^{ki}, \quad (7.10)$$

where  $C$  is an undetermined constant, and  $k$  may be 1, 2, or 3. Some of the cofactors  $E^{ki}$  may be zero, but there are always sufficient non-zero ones to give a solution for the vector components  $(\bar{\mathbf{a}}_1)_i$ . We repeat this process with the other two roots, and obtain the components of two other unrotated vectors,  $\bar{\mathbf{a}}_2$  and  $\bar{\mathbf{a}}_3$ . The vectors  $\bar{\mathbf{a}}_i$  are called the eigenvectors of  $\mathbf{S}$ , and may be used as a new basis  $\bar{\mathbf{A}}$ . If the transformation from the new basis to the old is represented by  $\mathbf{J} = (\mathbf{A} \mathbf{J} \bar{\mathbf{A}})$ , the matrix  $\mathbf{J}$  has columns  $J_{ji}$  consisting of the components  $(\bar{\mathbf{a}}_i)_j$  of the eigenvectors  $\bar{\mathbf{a}}_i$  (see p. 32). The deformation is represented in the new basis by the matrix  $\bar{\mathbf{S}} = (\bar{\mathbf{A}} \mathbf{S} \bar{\mathbf{A}})$ , and from (7.4)

$$\bar{\mathbf{S}} = (\bar{\mathbf{A}} \mathbf{J} \mathbf{A}) (\mathbf{A} \mathbf{S} \mathbf{A}) (\mathbf{A} \mathbf{J} \bar{\mathbf{A}}) = \mathbf{J}^{-1} \mathbf{S} \mathbf{J}.$$

From (7.9)

$$(\mathbf{S} \mathbf{J})_{ik} = S_{ij} J_{jk} = (S_{ij} - \lambda_k \delta_{ij}) J_{jk} + \lambda_k \delta_{ij} J_{jk} = \lambda_k J_{ik},$$

so that

$$\mathbf{J}^{-1} \mathbf{S} \mathbf{J} = \lambda_k \delta_{jk}. \quad (7.11)$$

The matrix  $\bar{\mathbf{S}}$  is thus a diagonal representation of the deformation, as concluded above. The quantities  $\lambda_i$  give the ratios of the lengths of the vectors  $\bar{\mathbf{a}}_i$  after the deformation to their lengths before deformation.

We have seen that the components of  $\mathbf{J}$  are not determined absolutely, since each eigenvector contains an arbitrary constant. The diagonal matrix  $\bar{\mathbf{S}}$  represents the strain  $\mathbf{S}$  in all coordinate systems with axes parallel to  $\bar{\mathbf{a}}_i$ , and the magnitudes of the measure lengths may be chosen arbitrarily. There is also an arbitrary choice of the order of the columns of  $\mathbf{J}$ , and correspondingly of the elements of  $\bar{\mathbf{S}}$ , since we may label any root of (7.7) as  $\lambda_1$ . This arises because we are free to label our coordinate axes in the basis  $\bar{\mathbf{A}}$  in any way we please, giving six different axis transformations from  $\mathbf{A}$  to  $\bar{\mathbf{A}}$ . We shall return to this question later, when discussing the idea of correspondence between directions and planes in different lattices.

When there are three real roots of (7.7), the unrotated directions may be used to specify three unrotated planes. Alternatively, these may be obtained by considering the condition that the vectors  $\mathbf{h}$  and  $\mathbf{k}$  which represent a plane normal before and after deformation are parallel. If  $\mathbf{h}$  and  $\mathbf{k}$  are the representations of  $\mathbf{h}$  and  $\mathbf{k}$  as column matrices in the bases  $\mathbf{A}^*$  (as before) we have

$$\mathbf{k}' = \mathbf{h}' \mathbf{S}^{-1} = (1/\lambda_i) \mathbf{h}', \quad (7.12)$$

which corresponds to (7.6) and has non-trivial solutions only when (7.7) is satisfied. The vectors obtained by substituting the roots  $\lambda_i$  into (7.12) give the directions of the unrotated plane normals, and are reciprocal to the eigenvectors  $\bar{\mathbf{a}}_i$ . The quantities  $\lambda_i$  give the ratios of the initial to the final spacings of the unrotated planes.

Note that if any  $\lambda_i = 1$ , the corresponding eigenvector is an invariant line, i.e. a direction which is both unrotated and undistorted, and there is correspondingly a plane with an invariant normal. If there are two invariant lines, they define an invariant plane, and the basis  $\bar{\mathbf{A}}$  is no longer unique. The normal to an invariant plane is necessarily unrotated, but need not itself be invariant.

A special case of homogeneous deformation arises when there are three orthogonal directions which are unrotated by the deformation. The axes of the strain ellipsoid and the reciprocal strain ellipsoid then coincide, and the deformation is said to be a pure strain. The formal definition of a pure strain is that it is a deformation such that the three orthogonal directions which remain orthogonal retain their directions and senses.

Let us now consider a deformation  $\mathbf{v} = \mathbf{P}\mathbf{u}$  which is represented in an orthonormal basis  $\mathbf{l}$  by the equation

$$[\mathbf{l}; \mathbf{v}] = [\mathbf{lP l}] [\mathbf{l}; \mathbf{u}]. \quad (7.13)$$

Consider a second orthonormal basis  $\mathbf{k}$ , related to  $\mathbf{l}$  by

$$[\mathbf{k}; \mathbf{u}] = (\mathbf{k L l}) [\mathbf{l}; \mathbf{u}]. \quad (7.14)$$

The matrix  $(\mathbf{k L l}) \equiv \mathbf{L}$  represents an axis transformation which is merely equivalent to rotating the basis  $\mathbf{l}$  into a new position.<sup>†</sup> The components  $L_{ij}$  are thus the cosines of the angles between  $\mathbf{k}_i$  and  $\mathbf{l}_j$ . By writing  $\mathbf{i}_j = L_{ij}\mathbf{k}_i$ , we obtain

$$\mathbf{i}_j \cdot \mathbf{i}_k = L_{ij}L_{ik}\mathbf{k}_i \cdot \mathbf{k}_i = L_{ij}L_{ik}\delta_{ii} = L_{ij}L_{ik}$$

and similarly  $\mathbf{k}_j \cdot \mathbf{k}_k = L_{ji}L_{ki}$ . Since both these scalar products are equal to  $\delta_{jk}$ , we obtain the well known relations between the direction cosines of the axes

$$L_{ij}L_{ik} = L_{ji}L_{ki} = \delta_{jk},$$

or in matrix form

$$\mathbf{L L}' = \mathbf{l}, \quad \mathbf{L}' = \mathbf{L}^{-1}, \quad |\mathbf{L}| = \pm 1. \quad (7.15)$$

Such a matrix is called orthogonal. When  $|\mathbf{L}| = +1$ , the bases  $\mathbf{l}$  and  $\mathbf{k}$  both correspond to right-handed (or left-handed) sets of base vectors, and the transformation of axes represented by  $\mathbf{L}$  is a proper rotation. When  $|\mathbf{L}| = -1$ , a right-handed set of base vectors is converted into a left-handed set, and vice versa; this is equivalent to a rotation plus a reflection in some plane, and is called an improper rotation. We consider only proper rotations.

If the strain (7.13) is now referred to the basis  $\mathbf{k}$ , its representation is given by (7.4)

$$(\mathbf{k P k}) = (\mathbf{k L l})(\mathbf{l P l})(\mathbf{l L k}) = (\mathbf{k L l})(\mathbf{l P l})(\mathbf{l L}' \mathbf{k}).$$

Hence taking the transpose of both sides

$$(\mathbf{k P}' \mathbf{k}) = (\mathbf{k L l})(\mathbf{l P}' \mathbf{l})(\mathbf{l L}' \mathbf{k}),$$

and it follows that if  $(\mathbf{l P l})$  is a symmetric matrix, so also is  $(\mathbf{k P k})$ . Symmetric matrices thus remain symmetric as a result of an orthogonal transformation.

If the deformation is a pure strain, we may choose the vectors  $\mathbf{k}_i$  to lie along the principal axes of this strain (i.e.  $\mathbf{k} = \bar{\mathbf{l}}$ ). The matrix  $(\mathbf{k P k}) = \bar{\mathbf{P}}$  is then diagonal, and  $(\mathbf{l P l})$  must therefore be symmetrical. A pure strain is thus characterized by a symmetric representation of the tensor  $\mathbf{P}$  in any orthonormal basis.

<sup>†</sup> We use the symbol  $\mathbf{L}$  for axis transformations which are pure rotations, and  $\mathbf{J}$  for more general axis transformations. A rotation contains three independent quantities, two to specify the axis of rotation and one to fix the magnitude of the rotation about this axis. It may also be described by a Rodrigues vector  $\mathbf{r} \tan(\theta/2)$  used in a Rodrigues-Frank map (Frank, 1988) or by a quaternion  $[r_1 \sin \theta/2, r_2 \sin \theta/2, r_3 \sin \theta/2, \cos \theta/2]$  (Handscorn 1958).

A pure deformation is equivalent to simple extensions or contractions along the three principal axes of strain. If the material consists of a rectangular parallelepiped with edges along the principal axes, it will thus remain a rectangular parallelepiped after deformation. The ratio of the new volume to the original volume is then  $\lambda_1\lambda_2\lambda_3 = |\mathbf{P}|$ , where  $\mathbf{P}$  is the matrix representation of the strain in any orthonormal basis, and  $\lambda_i$  are the eigenvalues of  $\mathbf{P}$ .

In simple cases, the pure strain matrix  $\mathbf{P}$  may be reduced to diagonal form by inspection. Thus if the basis  $\mathbf{i}$  may be transformed into the basis  $\bar{\mathbf{i}}$  by a single rotation about one of the vectors  $\mathbf{i}_i$ , the components  $L_{ij}$  ( $i \neq j$ ) of the rotation matrix are zero. In the general case, the problem is equivalent to finding the principal axes of a quadric surface. We use the procedure of p. 44, and since  $\mathbf{P}$  is symmetric, the roots of the equation

$$(\mathbf{P} - \lambda_i \mathbf{I}) \mathbf{x} = 0, \tag{7.16}$$

are necessarily all real. This equation is identical with an equation known as the discriminating cubic of the geometrical surface with scalar equation

$$\mathbf{x}' \mathbf{P} \mathbf{x} = P_{ij} x_i x_j = \text{const.} \tag{7.17}$$

As before, the columns of the matrix  $\mathbf{J}$  which diagonalizes  $\mathbf{P}$  are multiplied by undetermined constants. However, we wish the basis  $\bar{\mathbf{i}}$  to be orthonormal, so we must normalize the eigenvectors by choosing the constants so that eqns. (7.15) are satisfied. In this way,  $\mathbf{J}$  becomes an orthogonal matrix  $\mathbf{L}$  which is unique, except for the order of its columns. Three of the six ways in which the columns can be arranged correspond to improper rotations; of the remaining three, the most obvious choice is to label the principal axis which makes the smallest angle with  $\mathbf{i}_1$  as  $\bar{\mathbf{i}}_1$ , etc.

A pure strain is one of the two component deformations into which any homogeneous deformation may be analysed. The other type is a pure rotation, characterized by the condition that all vectors remain the same length. Obviously, a pure rotation is given by a tensor relation  $\mathbf{v} = \mathbf{R}\mathbf{u}$ , in which the components of the tensor  $\mathbf{R}$  form an orthogonal matrix in an orthonormal basis. This follows since if  $\mathbf{v}$  and  $\mathbf{u}$  are the vector representations in the orthonormal basis

$$|\mathbf{v}|^2 = \mathbf{v}' \mathbf{v} = \mathbf{u}' \mathbf{R}' \mathbf{R} \mathbf{u} = \mathbf{u}' \mathbf{u} = |\mathbf{u}|^2,$$

provided  $\mathbf{R}$  is orthogonal.

Any homogeneous deformation may be regarded as the result of a pure strain combined with a pure rotation. Thus we may write  $\mathbf{v} = \mathbf{S}\mathbf{u}$  as

$$\mathbf{v} = \mathbf{S} \mathbf{u} = \mathbf{P}_1 \mathbf{R}_1 \mathbf{u} = \mathbf{R} \mathbf{P} \mathbf{u},$$

where  $\mathbf{P}_1$ ,  $\mathbf{P}$  represent pure strains, and  $\mathbf{R}_1$ ,  $\mathbf{R}$  pure rotations.

Note that  $\mathbf{P}_1 \neq \mathbf{P}$  and  $\mathbf{R}_1 \neq \mathbf{R}$ ; there are two ways of resolving the deformation, depending on whether the rotation or the pure strain is considered to occur (mathematically) first. For the present, we find it convenient to use the second resolution, in which:

$$\mathbf{y} = \mathbf{R} \mathbf{P} \mathbf{x}, \tag{7.18}$$

means that the vector  $\mathbf{x}$  is first given a pure strain  $\mathbf{P}$  and then a pure rotation  $\mathbf{R}$  to form a new vector  $\mathbf{y}$ .

We have already seen that the ratio of the transformed volume to the original volume during the pure strain is given by  $|\mathbf{P}|$ . Since the rotation cannot change the volume, this quantity is also equal to the volume ratio for the whole deformation  $\mathbf{S}$ . Moreover, since  $|\mathbf{R}| = 1$ ,  $|\mathbf{S}| = |\mathbf{P}|$ , so that for any affine transformation, the volume ratio is given by  $|\mathbf{S}|$ . Finally, we note from eqn. (7.4) that for any axis transformation,  $|\mathbf{S}^B| = |\mathbf{S}|$ . The volume ratio is thus given by the determinant of *any* matrix representation of  $\mathbf{S}$ .

The geometrical relations involved in the general deformation may be appreciated by reference to Fig. 2.2, which is, however, two-dimensional. As a result of the deformation  $\mathbf{S}$ ,

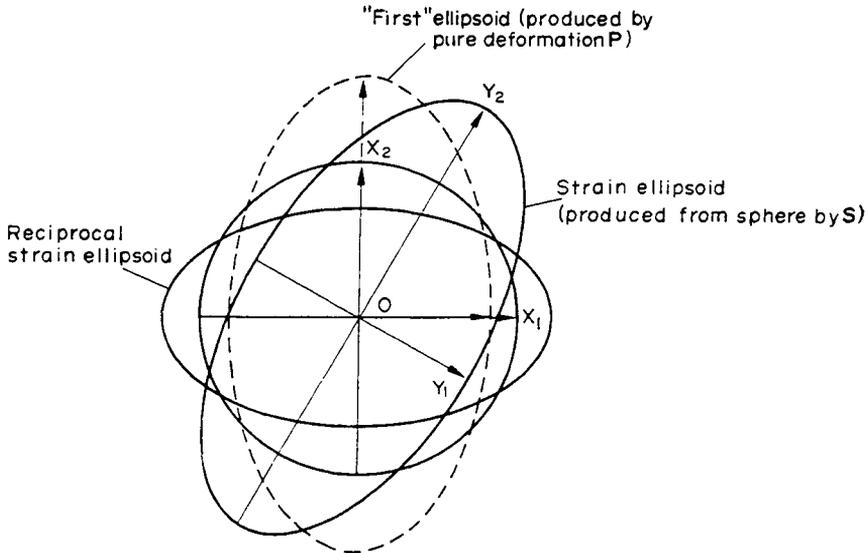


FIG. 2.2. Geometrical relations in homogeneous deformation.

The axis about which the first ellipsoid is rotated into the strain ellipsoid may have any orientation, but is assumed to be  $OX_3$  in order to give a two-dimensional figure.

a sphere is distorted into the strain ellipsoid, and the vectors  $OX_1$ ,  $OX_2$  (and  $OX_3$  not shown) become  $OY_1$ , etc. The deformation is regarded as taking place in two parts; during the pure strain, the vectors  $OX_1$ , etc. undergo simple extension or contraction to produce an ellipsoid which is shown dotted. There is no standard name for this figure, but we shall refer to it as the *first* ellipsoid. Following this, there is a rotation (not necessarily in the plane of  $OX_1$  and  $OX_2$ ) to the positions  $OY_1$ , etc.

The reciprocal strain ellipsoid will obviously be changed into a sphere by the pure strain  $\mathbf{P}$ , since  $\mathbf{R}$  produces no change in shape. The axes of this ellipsoid will thus be coincident with the principal axes of strain, as previously assumed. Given the matrix representation of the general deformation, in any set of orthonormal axes, we may resolve into  $\mathbf{P}$  and  $\mathbf{R}$  as follows. Any vector  $\mathbf{x}$  is converted into  $\mathbf{y} = \mathbf{S}\mathbf{x}$ . Now suppose  $\mathbf{x}$  represents a radius vector

of the reciprocal strain ellipsoid. Then after deformation, the components  $y$  satisfy the equation

$$y_1^2 + y_2^2 + y_3^2 = \text{const.}$$

or in matrix form

$$y' y = \text{const.}$$

Substituting for  $y$  in terms of the original components  $x$ , we obtain the scalar equation of the reciprocal strain ellipsoid

$$x' S' S x = S_{ki} S_{kj} x_i x_j = \text{const.} \quad (7.19)$$

By expanding  $S$  into its components  $R$  and  $P$

$$S' S = P' R' R P = P' P = P^2 \quad (7.20)$$

and this set of equations is sufficient to determine  $P$ . The symmetric matrix  $S' S$  has eigenvalues  $\lambda_i^2$  which are the squares of the eigenvalues of  $P$ . The reciprocal strain ellipsoid has the property that the ratio of the length of any deformed vector to its original length is proportional to the inverse radius vector of the ellipsoid drawn in the original direction.

From the eigenvalues of  $S' S$ , we can construct the matrix  $\bar{P}$  which is the diagonal representation of  $P$  in an orthonormal system along the principal axes, and we can also find the orthogonal transformation  $L = (\bar{I} L I)$  which transforms the basis  $I$  into the principal basis  $\bar{I}$ . The representations of the components of  $S$  in the original basis are then given by

$$\text{and} \quad \left. \begin{aligned} P &= L^{-1} \bar{P} L \\ R &= S P^{-1}. \end{aligned} \right\} \quad (7.21)$$

When referred to principal axes, the whole deformation takes the form

$$\bar{y} = \bar{S} \bar{x} = L R L^{-1} \bar{P} \bar{x}. \quad (7.22)$$

We can also show that the above procedure gives the principal axes without explicit reference to the reciprocal strain ellipsoid. Suppose we have two vectors,  $x_1$  and  $x_2$ , which are converted into two orthogonal vectors,  $y_1$ , and  $y_2$ . Then in the basis  $I$

$$y_1 \cdot y_2 = y_1' y_2 = x_1' S' S x_2 = 0.$$

Now if the two vectors were perpendicular before the deformation,  $x_1' x_2 = 0$ , so that

$$x_1' S' S x_2 = \lambda_i^2 x_1' x_2.$$

where  $\lambda_i$  is a scalar. The vectors  $x_1$  and  $x_2$  are then both solutions of the equation

$$(S' S - \lambda_i^2 I) x = 0 \quad (7.23)$$

and for non-trivial solutions

$$|S' S - \lambda_i^2 I| = 0. \quad (7.24)$$

This gives three orthogonal vectors which define the principal axes, and the procedure is equivalent to diagonalizing the matrix  $S' S$ , as described above. Moreover, if the three roots

are  $\lambda_1^2, \lambda_2^2, \lambda_3^2$ , we have

$$|\mathbf{y}_1|^2 = \mathbf{y}'_1 \mathbf{y}_1 = \lambda_1^2 \mathbf{x}'_1 \mathbf{x}_1,$$

so that the principal deformations are  $\lambda_i$ .

The above equations have to be modified when the deformation is expressed in a general basis  $A$ . The scalar product of  $\mathbf{y}_1$  and  $\mathbf{y}_2$  is now written

$$\mathbf{y}_1 \cdot \mathbf{y}_2 = (\mathbf{y}_1; A) (A^* G A) [A; \mathbf{y}_2] = (\mathbf{x}_1; A) (A S A)' (A^* G A) (A S A) [A; \mathbf{x}_2]$$

so that the general equation corresponding to (7.23) is

$$\{(A S A)' (A^* G A) (A S A) - \lambda_i^2 (A^* G A)\} [A; \mathbf{x}] = 0, \quad (7.25)$$

and there is an obvious corresponding equation for finding the characteristic roots.

The strain ellipsoid has a scalar equation which may be found by making  $|\mathbf{x}|^2$  constant. Thus in the basis  $\mathbf{i}$

$$\mathbf{y}' (S^{-1})' S^{-1} \mathbf{y} = \text{const.} \quad (7.26)$$

This surface has the geometrical interpretation that the ratio of the length of a deformed vector to its original length is proportional to the radius vector of the ellipsoid drawn in the final direction of the vector.

Finally, we note that the surface

$$\mathbf{x}' (S - I) \mathbf{x} = \text{const.}, \quad (7.27)$$

is called the elongation quadric, or in linear elasticity theory (Section 10), the strain quadric. The surface may be either an ellipsoid or a hyperboloid; it has axes in the same directions as those given by eqn. (7.17), and in general these do not coincide with the principal axes of strain. The elongation quadric has the geometrical interpretation that the *extension* of any line, resolved in the original direction of that line, is inversely proportional to the square of the radius vector to the surface, drawn in the original direction of the line.

## 8. TWIN CRYSTALS

Solid metals are usually composed of a compact mass of separate crystals or grains, joined along arbitrary internal surfaces, and randomly orientated with respect to each other. The orientation relation between any two grains having the same crystal structure may be expressed by a tensor relation representing a pure rotation, and the transformation may always be achieved by a proper rotation, with  $|R| = +1$ . In crystals of fairly high symmetry, the relation may also be expressed as an improper rotation, if so desired. In certain crystals which possess no centre of symmetry and few planes of symmetry, an improper rotation may produce an atomic arrangement which is not obtainable by a proper rotation. Such arrangements are called optical isomorphs, and show optical activity, i.e. the ability to rotate the plane of polarized light. They do not occur in metals.

The relation between two randomly orientated crystals thus requires three degrees of freedom for its specification, since there are three independent quantities in a rotation mat-

rix. If these crystals meet along a grain boundary surface, two further parameters are needed to specify the orientation of this surface at any point. The general grain boundary thus has five degrees of freedom. The concept of grain boundaries really belongs to the subject of crystal imperfections, and is considered in Chapter VIII. We may usefully consider here, however, the transformation between two orientations which are related in a well specified manner to the symmetry of the structure, so that the crystals are said to be twins of each other. Two crystals in twinned orientation may still be joined along any surface, but there is always some plane which will give a boundary of very low energy, and this composition plane then has no degrees of freedom.

Two crystals are twins of each other when they may be brought into coincidence either by a rotation of  $180^\circ$  about some axis (the twinning axis) or by reflection across some plane (the twinning plane).<sup>†</sup> The possible orientation relations may be further classified by the relation of the axis of symmetry to the composition plane of low energy along which the twin crystals are usually joined. The rotation axis may be normal to this composition plane (normal twins) or parallel to the composition plane (parallel twins). If the atomic structure is centrosymmetric, it follows that a normal rotation twin is equivalent to a reflection in the composition plane, which is then the twinning plane. For non-centrosymmetric structures, the operations of normal rotation and reflection will produce different twins having the same composition plane. Similarly, a parallel rotation twin is equivalent to a twin produced by a reflection in a plane normal to the rotation axis, and hence to the composition plane, for centrosymmetric structures, but these two operations produce different twins in non-centrosymmetric structures. There are thus four possible types of orientation relation between two twins, reducing to two equivalent pairs for centrosymmetric structures (and, of course, lattices).

Since most metallic structures are centrosymmetric, articles on twinning sometimes refer only to two types of twin orientation. These are then designated as "reflection" twins (equivalent to normal rotation twins) and "rotation" twins (equivalent to reflection in a plane normal to the composition plane). We have presented these results in axiomatic form, but they follow naturally from the condition that the two lattices fit together exactly along the composition plane. We shall emphasize this aspect of twinning in a more complete treatment in a later chapter.

Obviously, it is not possible to have reflection twins in which the twinning plane is a plane of symmetry in the crystal structure, since the twinning operation then merely reproduces the original orientation. In the same way, the twinning axis in a rotation twin can never be an even axis of symmetry. When two crystals are joined in twin orientation, the twinning plane and/or the twinning axis become pseudo-symmetry elements of the composite structure.

Most metallic crystals form twins which may be regarded both as reflection and normal rotation twins, though there are other types in metals of low symmetry. We could specify the twin relations by the rotation matrix  $R$  required to bring the crystals into coincidence

<sup>†</sup> A more general description of a mechanical twin (Crocker, 1962) is any region of the parent which has undergone a homogeneous shear to give a re-orientated region with the same crystal structure. The above orientation relations are then not necessarily valid; this is discussed in Chapter 20.

with each other, but it is often more useful to employ another type of deformation tensor. In many metals, and some non-metallic crystals, twins may be formed by a physical deformation of the original structure, and this is known as glide twinning. The process is macroscopically equivalent to a homogeneous shear of the original structure, and we therefore use this kind of deformation to describe the twinning law.

Figure 2.3 shows a section through a three-dimensional lattice which has undergone glide twinning. The open circles represent the lattice points in their original positions; the

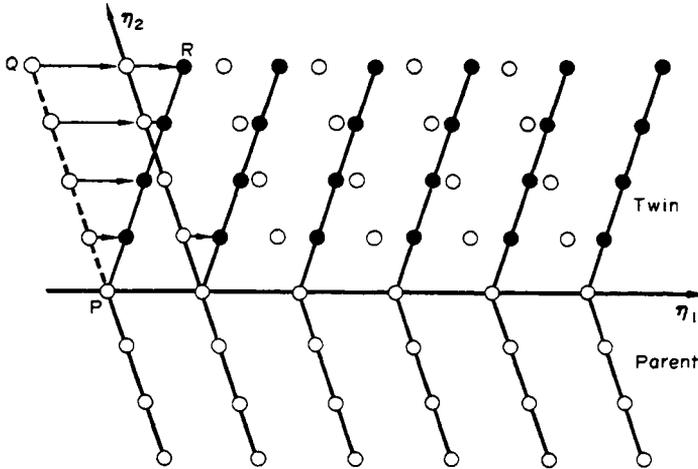


FIG. 2.3. Simple illustration of glide twinning.  
 $PQ$ ,  $PR$  are corresponding directions in parent and twin.

filled circles are the final positions to which they move. It will be seen that, in the twinned region, each lattice point moves in the same direction through a distance proportional to its distance from the composition plane, which is parallel to the direction of movement. In the simple example shown in the figure, the twinned structure is a reflection of the original structure in the composition plane, and the twinned lattice is obtained by a homogeneous simple shear of the original lattice. The composition, or twinning, plane is conventionally denoted  $K_1$  and the direction of shear  $\eta_1$ . The plane containing  $\eta_1$  and perpendicular to  $K_1$  (i.e. the plane of the diagram) is called the plane of shear.<sup>†</sup> Note that each lattice point moves only a fraction of the lattice repeat distance.

In the last paragraph, we emphasized that the two *lattices* in glide twinning are related by a homogeneous shear. The two *structures* are not so related unless the primitive unit cell contains only one atom, i.e. the structure may be obtained by placing an atom on each point of the lattice. More generally, some of the atoms must move in different directions from the lattice points; inhomogeneous movements of this kind are sometimes called atomic "shuffles". The *macroscopic* effect of the deformation is unaffected by the shuffles.

<sup>†</sup> The  $K_1$  plane is sometimes referred to as the shearing plane. This usage is better avoided, because of the possibility of confusion between "shearing plane" and "plane of shear".

We can thus describe the twin orientation by a matrix representing a simple shear, providing we confine our attention to the lattice and ignore the vectors  $\xi_{A,n}$  of eqn. (5.8).

Figure 2.4 shows the section of an original sphere which becomes an ellipsoid after the deformation. The section is in the plane of shear; vectors perpendicular to this plane are unaffected by the deformation, so the problem is essentially two-dimensional. If we use an orthonormal basis, with axes  $x_i$  parallel to  $\eta_1$ , perpendicular to the plane of shear and perpendicular to  $K_1$  respectively, we may specify the twin relation as

$$y = Sx,$$

where

$$S \equiv \begin{pmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (8.1)$$

and  $s$ , the amount of shear, is the distance moved by a lattice point at unit distance from the plane  $K_1$ . We note that  $|S| = 1$ , so there is no volume change in the transformation, as is physically obvious.

From Fig. 2.4., we see that the original sphere and the strain ellipsoid meet in two circ-

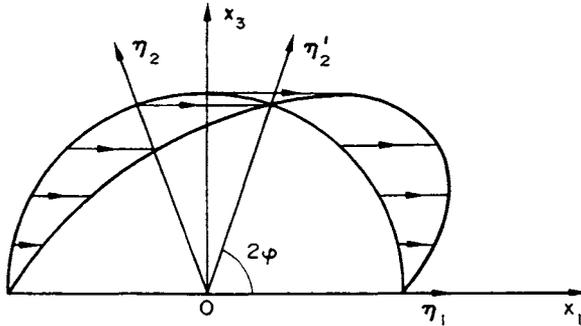


FIG. 2.4. Geometry of glide twinning.  
 $s$  is the displacement at unit distance along  $Ox_3$ .

les, which thus represent planes which are undistorted by the transformation. One of these is the composition plane  $K_1$  and this is also unrotated; the second plane is rotated through an angle  $\pi - 4\varphi = 2 \arctan(s/2)$  where  $2\varphi$  is the angle between the undistorted planes in their final positions. The second undistorted plane is denoted  $K_2$ , and its intersection with the plane of shear is in the direction  $\eta_2$ . The relation between  $s$  and  $\varphi$  is

$$s = 2 \cot 2\varphi \quad (8.2)$$

and the equation of the  $K_2$  plane in its initial and final positions is

$$x_1/x_3 = \mp s/2. \quad (8.3)$$

The matrix  $S$  is not symmetrical, and the deformation is thus not a pure shear. Using the method of pp. 44-7 or more simply from the geometry of Fig. 2.4, we find that the prin-

principal axes of strain are obtained by a right-handed rotation of  $x_i$  through an angle  $\varphi$  about  $x_2$ . Thus maximum extension takes place in a direction at  $\pi/2 - \varphi = \arctan\left[\frac{1}{2}\{s + (s^2 + 4)^{1/2}\}\right]$  to  $\eta_1$ , and the change in length is in the ratio  $\left[1 + \frac{1}{2}s\{s + (s^2 + 4)^{1/2}\}\right]^{1/2} : 1$ . Maximum contraction takes place in a direction at an angle  $\varphi = \arctan\left[\frac{1}{2}\{s - (s^2 + 4)^{1/2}\}\right]$  to  $\eta_1$ , and the change in length is in the ratio  $\left[1 + \frac{1}{2}s\{s - (s^2 + 4)^{1/2}\}\right]^{1/2} : 1$ .

The simple shear is equivalent to a pure shear, specified by the above extension and contraction, followed by a right-handed rotation of  $\pi/2 - 2\varphi$  about  $x_2$ . During the pure shear, the  $K_1$  and  $K_2$  planes rotate through equal angles  $\pi/2 - 2\varphi$  in opposite directions. It is thus clear that if we superimpose a left-handed rotation of this amount on the pure shear, the plane hitherto denoted  $K_2$  will be unrotated in the total deformation. The role of the  $K_1$  and  $K_2$  planes, and also of the  $\eta_1$  and  $\eta_2$  directions is thus interchanged; twins so related are called reciprocal or conjugate twins. Two crystals in twin orientation may be transformed into reciprocal twin orientation by a relative rotation of  $\pi - 4\varphi = 2 \arctan(s/2)$  about the normal to the plane of shear.

The above specification of glide twinning does not yet include the most important condition, namely that the twinning deformation should produce an equivalent lattice to the original lattice. If the twin is a mirror image in the composition plane, it is geometrically obvious that lattice points which are reflections of each other after twinning must have been separated by a vector parallel to  $\eta_2$  before twinning. For this type of twinning, therefore,  $\eta_2$  must be a rational direction parallel to a row of lattice points. The lattice structure will obviously be preserved if any three non-coplanar lattice vectors in the parent crystal are transformed into vectors in the twin which retain their lengths and mutual inclinations. All vectors which remain unchanged in length lie in either  $K_1$  or  $K_2$ , so we must select one lattice vector from one of these planes, and two from the other. This means that either  $K_1$  or  $K_2$  must be a rational plane. The angle between any two vectors in either  $K_1$  or  $K_2$  is unchanged by the twinning shear, but in general a vector in  $K_1$  and a vector in  $K_2$  change their relative inclination. However, the angle between  $\eta_1$  and any vector in  $K_2$  is unchanged, as is the angle between  $\eta_2$  and any vector in  $K_1$ . Our three vectors may thus be  $\eta_2$  and any two lattice vectors in  $K_1$ , or  $\eta_1$  and any two lattice vectors in  $K_2$ . In the first case, we have  $\eta_2$  and  $K_1$  rational, and this is called a twin of the first kind; a twin of the second kind has  $\eta_1$  and  $K_2$  rational.<sup>†</sup>

The two possibilities are illustrated in Figs. 2.5 and 2.6. Twins of the first kind are a simple reflection of the parent crystal in the  $K_1$  plane, and the twin *lattice* may equivalently be obtained by a rotation of  $180^\circ$  about the normal to  $K_1$ . As mentioned above, these two descriptions of the *structure* are also equivalent if this is centrosymmetric. Twins of the second kind are related by a rotation of  $180^\circ$  about  $\eta_1$ , and for centrosymmetric structures, this is equivalent to a mirror reflection in the plane perpendicular to  $\eta_1$ , as shown in the figure.

<sup>†</sup> If the more general definition of mechanical twinning mentioned on p. 52 is adopted, twins with all four elements irrational are possible. In practice, such twinning modes probably result only from 'double twinning', that is, from the combination of two twinning operations of the type discussed in the text. More rigorous proofs of the statements in the text are given in Chapter 20.

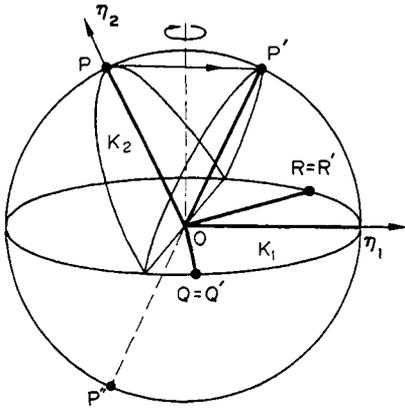


FIG. 2.5. Type I twinning.

$P, Q, R$  represent lattice vectors before twinning.  $P', Q', R'$  represent the corresponding vectors after twinning (after Cahn, 1953).

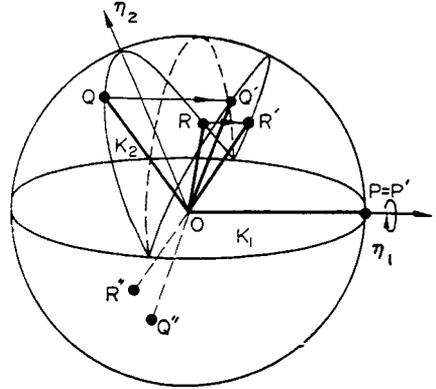


FIG. 2.6. Type II twinning.

$P, Q, R$  represent lattice vectors before twinning.  $P', Q', R'$  represent the corresponding vectors after twinning (after Cahn, 1953).

If the basic vectors used to define the twinning elements are all primitive vectors of the lattice, the homogeneous deformation will produce the twinned lattice, as assumed above. However, it is possible to produce twinned structures even when all these vectors are not primitive. In this case, some of the atomic movements must be inhomogeneous, even in a structure without a basis, since only a fraction of the lattice points move to their twinned positions as the result of the shear. In effect, the macroscopic shear now converts a superlattice of the structure into its twinned orientation.

For any crystal structure, the twinning elements and the amount of shear are completely specified by either  $\eta_2$  and  $K_1$  or by  $\eta_1$  and  $K_2$ . All four elements may be rational, and we then have compound twins. Most twins are compound in metals of cubic, trigonal and tetragonal symmetry, but both type I and type II twins have been found in orthorhombic  $\alpha$ -uranium. In metals of the highest symmetry,  $K_1$  and  $K_2$  and  $\eta_1$  and  $\eta_2$  are frequently crystallographically equivalent, so that the twin and its reciprocal represent equivalent twinning modes.

If the deformation is represented in a basis  $A$  of the parent crystal by  $S$ , we may now use the transformation formulae previously derived to find the new indices of any plane or lattice direction relative to  $A$ . If these new indices are  $k = [A^*; k]$  and  $v = [A; v]$  where  $k$  and  $v$  are the new plane normal and lattice vector, we have

$$k' = h' S^{-1} \quad \text{and} \quad v = S u.$$

In general, both directions and planes in the twin have irrational indices when referred to a basis in the parent. We then find it useful to introduce the idea of *correspondence* between the two lattices. Refer the vector  $v$  to a new basis which forms a unit cell in the twin. Since twin and parent have the same structure, it will be natural to choose a unit cell of the same size and shape as that outlined by the vectors  $a_i$ ; the new basis  $B$  will then differ from  $A$

only by a rotation (a proper rotation if both sets of base vectors are right-handed). We now have

$$\begin{aligned} \text{and } {}^B\mathbf{v} &= [\mathbf{B}; \mathbf{v}] = (\mathbf{B} \mathbf{L} \mathbf{A}) [\mathbf{A}; \mathbf{v}] \\ {}^B\mathbf{v} &= \mathbf{L} {}^A\mathbf{v} = \mathbf{L} \mathbf{S} {}^A\mathbf{u} = \mathbf{C} {}^A\mathbf{u}. \end{aligned} \quad (8.4)$$

The matrix  $\mathbf{C} = (\mathbf{B} \mathbf{C} \mathbf{A})$  is called the correspondence between the two lattices; it combines the effects of the deformation and the change of basis. Clearly, since  $\mathbf{v}$  is rational in  $\mathbf{B}$  and  $\mathbf{u}$  is rational in  $\mathbf{A}$ ,  ${}^B\mathbf{v}$  and  ${}^A\mathbf{u}$  both have rational components, and the components of  $\mathbf{C}$  are all rational. Moreover, since the bases  $\mathbf{A}$  and  $\mathbf{B}$  both refer to a unit cell of the same volume,  $|\mathbf{C}| = \pm 1$ . From the result on p. 45, it follows that the columns of  $(\mathbf{B} \mathbf{C} \mathbf{A})$  are the components referred to base  $\mathbf{B}$  of the vectors which are formed from the base vectors  $\mathbf{a}_i$  of  $\mathbf{A}$  by the transformation  $\mathbf{S}$ . If we use the transformed base vectors of  $\mathbf{A}$  as the base vectors of  $\mathbf{B}$  (symbolically,  $\mathbf{B} = \mathbf{C} \mathbf{A}$ ), the correspondence matrix is  $(\mathbf{C} \mathbf{A} \mathbf{C} \mathbf{A}) = \mathbf{I}$ . We can thus always establish a unitary correspondence between direction indices in the parent and in a suitable basis in the twin. In type I twinning, for example, we could choose the basis  $\mathbf{A}$  to have  $\mathbf{a}_1$  and  $\mathbf{a}_2$  in  $K_1$  and  $\mathbf{a}_3 = \mathbf{h} - \frac{1}{2}\mathbf{s}$  where  $\mathbf{h}$  is the vector normal to  $K_1$  and  $\mathbf{s}$  is a vector in  $K_1$  chosen so that  $\mathbf{a}_3$  is parallel to  $\eta_2$ . After twinning, the vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are unchanged and  $\mathbf{a}_3$  becomes  $\mathbf{h} + \frac{1}{2}\mathbf{s}$ . If these three vectors are used as the basis  $\mathbf{B}$ , then  $\mathbf{C} = \mathbf{I}$ . However, it is sometimes more convenient to choose bases so that the correspondence matrix is not the unit matrix. When this is done, the transformation of direction indices is given by eqn. (8.4), and the correspondence between plane indices is specified by

$$\left. \begin{aligned} (\mathbf{k}; \mathbf{B}^*) &= (\mathbf{k}; \mathbf{A}^*) (\mathbf{A}^* \mathbf{L} \mathbf{B}^*) \\ &= (\mathbf{h}; \mathbf{A}^*) (\mathbf{A} \mathbf{S}^{-1} \mathbf{A}) (\mathbf{A} \mathbf{L} \mathbf{B}) \\ &= (\mathbf{h}; \mathbf{E}^*) (\mathbf{A} \mathbf{C} \mathbf{B}) \end{aligned} \right\} \quad (8.5)$$

using the result of (6.11). This equation may be written more briefly as

$${}^B\mathbf{k}' = {}^A\mathbf{h}' \mathbf{C}^{-1}.$$

## 9. RELATIONS BETWEEN DIFFERENT LATTICES

In deformation twinning, the magnitude of the shear is fixed, once the twinning elements have been specified. An arbitrary shear would produce a new lattice with a different symmetry, but having the same volume. More generally, any space lattice may be converted into any other lattice by a homogeneous deformation  $\mathbf{S}$ . The deformation determines not only the symmetry and parameters of the new lattice, but also its orientation relative to the original lattice. The orientation of the interphase boundary (i.e. the surface of separation) must be specified separately.

When we have two crystal structures in contact with each other in a fixed relative orientation, it is often convenient to choose a deformation tensor  $\mathbf{S}$  to describe the relation between the lattices. We have already emphasized that whilst any two unit cells of the lattices may be connected by a suitable homogeneous deformation, the positions of the atoms cannot necessarily all be described in this way. If the primitive unit cells of the two structures

contain different numbers of atoms, the conversion of one structure into another involves a net loss or gain of lattice points. In such cases, it seems sensible to choose a tensor  $\mathbf{S}$  which relates unit cells containing the same number of atoms, and this is essential in transformations where there is a correspondence (see below).

In the case of twinning, we emphasized that shuffles may be produced either because the structure contains more than one atom per unit cell, or because the simple shear  $\mathbf{S}$  does not relate all the lattice points. The same conclusion applies to more general lattice deformations; the smallest unit cells related by  $\mathbf{S}$  need not be primitive cells of either structure, and shuffling is then required to complete the transformation even if both structures have only one atom per lattice point. A distinction is sometimes made between these latter type "lattice shuffles", which are determined by choice of  $\mathbf{S}$ , and the more general "structure shuffles", which arise from the atomic position vectors  $\xi$  of (5.8).

For complete generality, the relation between any two lattices should be written in the form

$$\mathbf{v} = \mathbf{t} + \mathbf{S}\mathbf{u}, \quad (9.1)$$

where the tensor  $\mathbf{S}$  specifies the relative sizes and orientations of the unit cells, and the constant vector  $\mathbf{t}$  represents a translation of the lattice points of one crystal relative to those of the other. In general, such translations are not detectable by ordinary crystallographic methods, and are of interest only when the actual atomic positions in two lattices separated by an interface are being considered (Section 36). A special case is when  $\mathbf{S} = \mathbf{I}$ ; the relation then describes a surface defect, known as a stacking fault, in a single lattice (Section 16). For the remainder of this section, we assume  $\mathbf{t} = 0$  and consider only the orientation and structural relations specified by  $\mathbf{S}$ .

In any structure, there is an infinite number of operations which will bring the lattice into self-coincidence. Correspondingly, there is an infinite number of deformations  $\mathbf{S}$  which will convert a specified unit cell into another specified unit cell with given orientation relation. It is clear that as long as we wish merely to give a formal statement of the relative positions of the two sets of lattice points, any deformation  $\mathbf{S}$  which gives the desired relation is valid. In some phase transformations, however, the atoms in a region of a product crystal have moved from their original positions in the parent crystal in such a way that the lattice of the parent has been effectively deformed into the lattice of the product. There is then a particular tensor  $\mathbf{S}$  which not only specifies the relations of the lattices, but also the way in which one lattice may change into another. This tensor has usually to be selected by some external physical assumption, e.g. that each point of the original lattice moves to the nearest point of the final lattice. The simplest example of the physical significance of the choice of  $\mathbf{S}$  occurs in twinning. A particular twinning law could be represented by a rotation about a suitable axis by any odd multiple of  $\pi$ , or alternatively by a simple shear deformation. For mechanical twinning, the latter statement of the law is more meaningful, since one lattice is physically sheared into the other.

A deformation which is physically significant implies a one to one correspondence between vectors in the two lattices. Each vector in one lattice may be associated unambiguously

with a "corresponding" vector of the other lattice into which it is converted by the transformation. We summarize these relationships by means of a correspondence matrix, as already used for twin crystals.

Suppose we use bases  $A$  and  $B$  in the two lattices which we call  $\alpha$  and  $\beta$  respectively; the unit cells defined by  $A$  and  $B$  need not contain the same numbers of atoms. The relation between vectors in the two lattices is given by

$$[A; \mathbf{v}] = (A S A) [A; \mathbf{u}],$$

and the relation between the two bases is

$$[B; \mathbf{u}] = (B J A) [A; \mathbf{u}].$$

Combining these two equations,

$$\begin{aligned} \text{or briefly} \quad [B; \mathbf{v}] &= (B J A) (A S A) [A; \mathbf{u}] = (B C A) [A; \mathbf{u}], \\ &B \mathbf{v} = C A \mathbf{u}. \end{aligned} \quad (9.2)$$

As on p. 56, the columns of the correspondence matrix  $C = (B C A)$  are the components referred to basis  $B$  of the vectors which are formed from the base vectors  $\mathbf{a}_i$  by the deformation  $S$ . If these transformed vectors are used to define the new unit cell ( $B = CA$ ), the correspondence matrix is  $I$ . However, this will not usually happen if  $B$  and  $A$  are derived from conventional unit cells in the two structures.

On pp. 47-8 we showed that the determinant of  $(A S A)$  gives the ratio of the volume of the  $\beta$  structure to that of the  $\alpha$  structure. If the bases  $A$  and  $B$  contain the same number of atoms, the determinant of  $(B J A)$  will give the ratio of the volume per atom in the  $\alpha$  structure to that in the  $\beta$  structure, and hence the determinant of  $(B C A)$  will be unity.<sup>†</sup> Correspondingly, if there are different numbers of atoms in the two bases (as may often happen if  $\mathbf{a}_i$  and  $\mathbf{b}_i$  define primitive or conventional unit cells),  $|C|$  will equal the ratio of the number of atoms in the unit cell defined by  $A$  to the number in the unit cell defined by  $B$ . The elements of  $(B C A)$  will all be rational; that is, they are small integers or fractions.

The reverse transformation is clearly specified by

$$[A; \mathbf{u}] = (A C B) (B; \mathbf{v}),$$

where  $(A C B)$  is the reciprocal matrix  $(B C A)^{-1}$ .

The correspondence of directions also implies a one to one correspondence of lattice planes, since from eqns. (7.5) and (6.10)

$$(\mathbf{k}; A^*) = (\mathbf{h}; A^*) (A S A)^{-1},$$

and

$$(\mathbf{k}; B^*) = (\mathbf{h}; A^*) (A S A)^{-1} (A J B) = (\mathbf{h}; A^*) (A C B), \quad (9.3)$$

or briefly

$$\mathbf{k}' = \mathbf{h}' C^{-1}.$$

<sup>†</sup> The determinant  $|C|$  is  $\pm 1$  depending on whether or not both bases are defined by equal-handed sets of vectors.

In simple shear, the lattice points move parallel to the  $K_1$  plane through distances which are proportional to their separation from this plane. This is an example of plane strain in which all displacements are coplanar. A more general type of homogeneous plane strain is shown in Fig. 2.7. Once again all lattice points move in the same direction through distances which are proportional to their separation from a fixed plane, but the direction of movement is no longer parallel to this plane. There are again two undistorted planes; one is also unrotated, and corresponds to the  $K_1$  plane in mechanical twinning. The whole

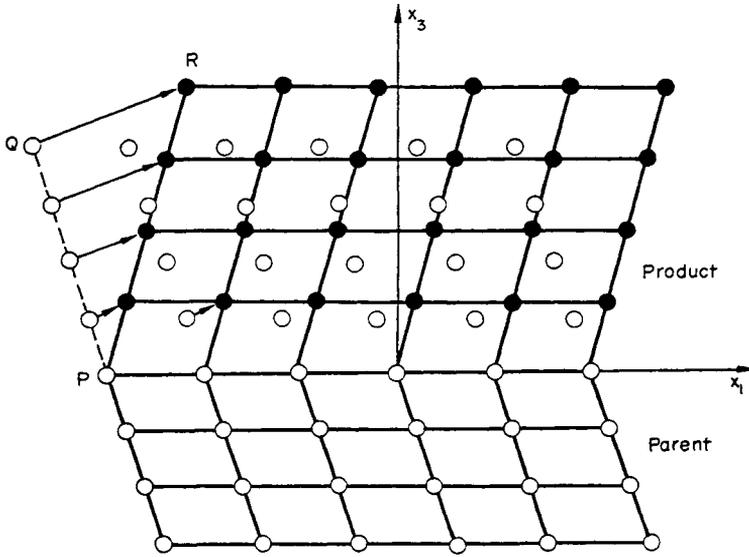


FIG. 2.7. Simple illustration of an invariant plane strain connecting two lattices.  $PQ, PR$  are corresponding directions in parent and product.

deformation may be considered as the relative displacement of a stack of such planes in a fixed direction.

The type of deformation shown in Fig. 2.7 is called an invariant plane strain. We can readily prove that the general condition for  $S$  to represent such a strain is that one principal strain be zero, and the other two have opposite signs. Thus in an orthonormal basis in which the deformation is  $y = Sx = RPx$ , we refer the representations to the principal axes of strain (basis  $\bar{i}$ ) and obtain (see eqn. (7.22))

$$\bar{y} = LRL\bar{P}\bar{x},$$

where  $\bar{P} = (\lambda_i\delta_{ij})$  is a diagonal matrix. A vector is unchanged in length if

$$\bar{y}'\bar{y} = \bar{x}'\bar{P}'\bar{P}\bar{x} = \bar{x}'\bar{x},$$

since  $L$  and  $R$  are both orthogonal. In scalar form, this equation is

$$(\lambda_i^2 - 1)\bar{x}_i^2 = 0, \tag{9.4}$$

and gives the locus of the surface containing lines of unchanging length. If there are two such lines which are coplanar, they define an undistorted plane. By considering the intersection of (9.4) with the plane  $\bar{x}_k = 0$ , we obtain the equations

$$\frac{\lambda_i^2 - 1}{\lambda_j^2 - 1} = -\frac{\bar{x}_j^2}{\bar{x}_i^2}. \quad (9.5)$$

For a real solution, each  $\lambda_i^2 - 1$  must differ in sign from each of the other two, because  $\bar{x}_i^2$  are necessarily positive. This is only possible if one of the  $\lambda_i^2 - 1$  is zero, and the other two have opposite signs. Note that in contrast to this, the only condition for a plane strain is that one  $\lambda_i$  should equal 1. The invariant plane strain is thus not the most general kind of plane strain.

Equation (9.5) now represents a straight line which is the intersection of the plane surface

$$(\lambda_1^2 - 1)\bar{x}_1^2 + (\lambda_3^2 - 1)\bar{x}_3^2 = 0, \quad \bar{x}_2 = \text{const}, \quad (9.6)$$

with the plane  $\bar{x}_2 = 0$ . This plane is a plane of zero distortion. During the pure deformation P, the undistorted plane will rotate through some angle, so an invariant plane strain is obtained by combining P with a rotation R which returns the undistorted plane to its original position. The total deformation S is thus determined, once P has been specified.

The tensor S has a particularly simple representation if we choose axes  $x_i$  in Fig. 2.7 to correspond to those previously used for simple shear in deformation twinning. The whole deformation is then equivalent to a simple shear on the  $x_1x_2$  plane, combined with a uniaxial expansion or contraction perpendicular to this plane. In this system the deformation is thus given by

$$S = \begin{pmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 + \Delta \end{pmatrix}, \quad (9.7)$$

where  $s$  and  $\Delta$  specify the shear and the expansion, and  $1 + \Delta$  is the volume ratio of the transformation.

More generally, we now find the form of the tensor S in any basis A. A vector  $u$  in the undistorted plane must satisfy the relation

$$S u = u. \quad (9.8)$$

The normal to the undistorted plane is  $v$  where  $v' u = 0$ , and after deformation the normal is given by

$$v' S^{-1} = \{1/(1 + \Delta)\} v', \quad (9.9)$$

where  $\Delta$  is a scalar.

Quite generally, we may write  $S = I + S^\circ$ , and since  $S^\circ u = 0$  it follows that  $S^\circ$  must have the form  $e v'$ , where  $e$  is any constant vector. The above relations are thus satisfied if

$$S = I + e v' \quad (9.10)$$

and this is the general representation of the invariant plane strain on the plane with normal  $\nu$ . In suffix form

$$S_{ij} = \delta_{ij} + e_i \nu_j. \quad (9.11)$$

We also see from (9.9) that the scalar product  $\nu' e = \Delta$ . By expansions of  $|S|$ , the volume ratio of the new phase to the old is found to be  $1 + \Delta$ ;  $\Delta$  is called the cubical dilatation, or simply the dilatation.

The reciprocal matrix to  $S$  is found similarly to be

$$S^{-1} = I - \frac{e \nu'}{1 + \Delta}. \quad (9.12)$$

Any point with coordinates given by the position vector  $u$  moves to a position  $Su$ , and its displacement is thus

$$S u - u = e \nu' u, \quad (9.13)$$

which is always in the direction  $e$ . The amount of displacement is proportional to the perpendicular distance from the plane  $\nu' u = 0$ .

Similarly, any normal  $h$  becomes a vector  $k$  where  $k' = h' S^{-1}$ , and the displacement of the end point of this vector is thus

$$h' \left( I - \frac{e \nu'}{1 + \Delta} \right) - h' = - \frac{h' e \nu'}{1 + \Delta}. \quad (9.14)$$

This is always in the direction  $\nu$ , and hence each plane normal rotates in a plane containing  $\nu$ .

Crystals of a new phase frequently form inside an existing solid phase in the form of flat plates. In nucleation and growth reactions, this shape is usually adopted to minimize the elastic strain energy due to the volume change in the transformation, and the boundary surface need not have special significance in any representation of the orientation relations. In martensitic transformations, however, we have emphasized that some choice of  $S$  has physical significance in defining the atom movements. By analogy with the situation in mechanical twinning, we might expect that when the correct choice is made,  $S$  has the form (9.10), and the plane  $\nu$  specifies the interfacial boundary or habit plane of the martensite crystals. However, although there are an infinite number of ways in which two lattices may be related by a tensor  $S$ , it is seldom possible to find a representation of type (9.10). This is because the lattice parameters of the two structures are determined mainly by short range interactions, and the condition (9.5) will be satisfied for two different structures only coincidentally. In fact, similar planes of almost identical atomic arrangement do occur in certain transformations between the closely related f.c.c. and h.c.p. structures and the interface plane is then rational in both lattices. Apart from these isolated examples, however, the interpretation of the martensite habit plane cannot be as simple as the interpretation of the  $K_1$  plane in mechanical twinning.

As previously described, a general homogeneous deformation is characterized by at least one and possibly three unrotated planes. Jaswon and Wheeler (1948) pointed out that if the

martensite lattice is produced from the parent lattice by a homogeneous deformation, the associated disturbance of the untransformed parent phase in the region round a martensite crystal will be very large, unless the boundary between the phases is unrotated. They therefore suggested that the habit plane is an unrotated plane, and should be found by diagonalizing  $S$  in accordance with the procedure on p. 44. However, these unrotated planes are distorted, and the rotation of rows of atoms within them would equally give rise to an extremely large strain energy.

In considering martensitic transformation, we are thus faced with the difficulty that finite homogeneous deformation of a macroscopic region of the parent crystal is only feasible if the boundary between the deformed and undeformed regions at any stage at least approximates to an invariant plane of the deformation. This requirement can be reconciled with the impossibility of relating the lattices by a homogeneous deformation of this kind only by assuming that a martensite plate is not produced physically by a *homogeneous* deformation of the parent lattice. It is now generally accepted that the atom movements during transformation are such that the deformation of the parent lattice into the martensite lattice is homogeneous only over a localized region. An adjacent region of the plate is formed by a different physical deformation, which also, of course, generates the martensite structure. Each of these deformations may be factorized into two components, of which one is common to both regions, and the other is opposite in the two regions, in the sense that the combined effect of these components is to produce no net change in shape or volume. The whole transformation thus consists of a homogeneous deformation, together with deformations which are locally homogeneous, but which have zero macroscopic effect.

The geometry and crystallography of martensite transformations is considered in detail in Part II, Chapter 22; the subject is introduced here only in order to show the importance of analysing lattice deformations  $S$  into component deformations. From the above description, we see that the component of the lattice deformation common to all regions of a plate must approximate closely to an invariant plane strain. The component with zero macroscopic effect must be such that there is no dilatation (i.e. the determinant of its matrix representation must be unity), since any volume change has to be identical in all regions, and would thus accumulate to produce a macroscopic effect. It is generally assumed that this component is also a plane strain, that is, in this case, a simple shear. The description of this component depends on the order in which the separate deformations are considered to be applied, and this emphasizes the purely mathematical character of the factorization. The terms "first" and "second" strain, which are often used for convenience, have no physical significance, and this is true of all factorizations of this type.

Two invariant plane strains applied successively will not give a resultant deformation with an invariant plane unless the plane  $\nu$  or the direction  $e$  (eqn. (9.10)) is common to both components. This may be seen by writing the components  $D$ ,  $T$  in subscript form

$$D_{ik} = \delta_{ik} + (e_1)_i (v_1)_k, \quad T_{kj} = \delta_{kj} + (e_2)_k (v_2)_j.$$

The product  $S_{ij} = D_{ik}T_{kj}$  can only be written in the form  $\delta_{ij} + (e_3)_i (v_3)_j$  if either  $e_1 = e_2 = e_3$  or  $v_1 = v_2 = v_3$ . There must, however, always be an invariant line in such a

transformation, since one line is common to the two invariant planes. Any deformation tensor which is to be factorized into two invariant plane strains must thus be an invariant line strain. For any deformation, there is at least one unrotated line, but this need not be unchanged in length, and an invariant line strain is not the most general form of homogeneous deformation.

The condition for two lattices to be related by an invariant line strain is simply that one principal strain either be zero, or have opposite sign from the other two. This follows immediately from eqn. (9.5) since we then have for any point on an unextended line

$$(\bar{x}_1)^2 = - \frac{(\lambda_2^2 - 1)(\bar{x}_2)^2 - (\lambda_3^2 - 1)(\bar{x}_3)^2}{(\lambda_1^2 - 1)}, \quad (9.15)$$

and if the two principal strains  $\lambda_2 - 1$ ,  $\lambda_3 - 1$  have the same sign,  $\lambda_1 - 1$  must have the opposite sign. Provided this condition is satisfied, eqn. (9.15) represents a cone of directions of unchanging length. By addition of a suitable rotation, any line in the cone may be returned to its original position, so that it is an invariant line of the whole deformation. The rotation only affects the orientations of the two lattices, so that the condition above is sufficient to ensure that two lattices can be related by an invariant line strain.

In discussing invariant plane and line strains, we have imposed no restrictions on the nature of these planes or lines, since (in contrast to the situation in twinning) we make no general assumptions about the relations of the lattice symmetries or constants. It follows that the invariant planes may be irrational.

### 10. INFINITESIMAL DEFORMATIONS

The deformations considered previously in this chapter have been homogeneous and finite. In this section we briefly consider the relation of the results to the theory of linear elasticity, which deals with infinitesimal inhomogeneous displacements of the atoms. Let us first consider an affine transformation. Throughout this section, we shall use an orthogonal basis  $\mathbf{i}$  in which the representation of a general position vector is  $\mathbf{x}$ . Since we shall not be concerned with the crystal lattice, we use the symbol  $\mathbf{x}$  rather than  $\mathbf{u}$  for the vector itself, and the deformation

$$\mathbf{y} = \mathbf{S}\mathbf{x}$$

is represented by the matrix equation

$$y_i = S_{ij} x_j,$$

in the basis  $\mathbf{i}$ . As a result of the deformation, a point with coordinates  $x_i$  moves to a position with coordinates  $y_i = S_{ij}x_j$ . We define the displacement vector  $\mathbf{w}$  of this point as the vector joining its initial and final positions. The displacement vector thus has components  $w_i = y_i - x_i$ , and is given by the equation

$$\left. \begin{aligned} \mathbf{w} &= \mathbf{S}\mathbf{x} - \mathbf{x}, \\ \mathbf{w} &= (\mathbf{S} - \mathbf{I})\mathbf{x} \\ \text{i.e.} \quad \text{or} \quad w_i &= (S_{ij} - \delta_{ij})x_j. \end{aligned} \right\} \quad (10.1)$$

Now consider a further small deformation with matrix representation  $z = T y$ . The total deformation is thus

$$z = T S x$$

or in suffix form,

$$z_i = T_{ik} S_{kj} x_j = [(T_{ik} - \delta_{ik})(S_{kj} - \delta_{kj}) + (S_{ij} + T_{ij}) - \delta_{ij}] x_j. \quad (10.2)$$

Now suppose the deformations  $S$  and  $T$  are both infinitesimal, in the sense that the elements of the displacement vectors  $(S-1)x$ ,  $(T-1)y$  are so small that the product of any term in the matrix  $(S-1)$  with any other term, or with any term in the matrix  $(T-1)$ , may be neglected. This means that the first term in the square brackets of eqn. (10.2) is zero. The total displacement is thus given by

$$w_i = z_i - x_i = (S_{ij} + T_{ij} - 2\delta_{ij}) x_j$$

or, in matrix form,

$$w = (S + T - 2I) x. \quad (10.3)$$

The components of the displacement vector for two successive strains may thus be obtained by adding the components of the vectors for the separate strains. This result is obviously only valid for small displacements, and is known as the principle of superposition. When it is applicable, the resultant displacement is independent of the order of the strains. Conversely, we may factorize any infinitesimal deformation into two separate and independent components.

In principle, the infinitesimal affine deformation applied to any lattice will produce a new lattice. However, when the displacements are small enough for the principle of superposition to be applied, each lattice point in the deformed structure can be associated clearly and unambiguously with its original position. It is then more useful to regard the deformed structure as a slightly imperfect (strained) version of the original lattice, rather than as a new lattice. This point of view is, moreover, essential when later in this section we consider inhomogeneous displacements. In their deformed positions, the lattice points then no longer constitute a space lattice in our former mathematical use of the term. They may still be regarded as forming a slightly imperfect lattice of the original type.

Consider the equation

$$w = (S - I) x = Q x, \quad (10.4)$$

where the components of  $Q$  are all infinitesimal. In general,  $Q$  will not be symmetric, since  $S$  need not represent a pure deformation. However, we may always write

$$\left. \begin{aligned} Q &= e + \omega, \\ e_{ij} &= \frac{1}{2}(Q_{ij} + Q_{ji}), \\ \omega_{ij} &= \frac{1}{2}(Q_{ij} - Q_{ji}). \end{aligned} \right\} \quad (10.5)$$

The components  $e_{ij}$  form a symmetric matrix  $e$  which is a representation of the strain tensor, whilst the components  $\omega_{ij}$  form an antisymmetric matrix  $\omega$ . We shall now show that for infinitesimal deformations, an antisymmetric matrix is the representation of a rigid-body rotation.

The condition that the length of a vector should remain unchanged as the result of a deformation  $R$  was shown on p. 47 to be

$$\mathbf{x}' R' R \mathbf{x} = \mathbf{x}' \mathbf{x},$$

i.e.

$$R_{ki} R_{kj} x_i x_j = \delta_{ij} x_i x_j.$$

Expanding  $R_{ki} R_{kj}$ , this gives

$$\delta_{ij} = (R_{ki} - \delta_{ki})(R_{kj} - \delta_{kj}) + R_{ji} + R_{ij} - \delta_{ij}. \quad (10.6)$$

Our previous assumption shows that the first term on the right may be neglected, so that

$$R_{ij} + R_{ji} = 2\delta_{ij}. \quad (10.7)$$

If the matrix  $R$  satisfies this condition, the displacement vector is given by

$$\mathbf{w} = (R - I) \mathbf{x},$$

where the matrix  $R - I$  is antisymmetric. An antisymmetric displacement matrix thus represents an equal rotation of all vectors about the origin of coordinates, i.e. a rigid-body rotation.

From the Principle of Superposition, the general displacement vector may now be analysed into two components. The change  $\mathbf{w}_1 = \boldsymbol{\omega} \mathbf{x}$  represents a rotation, whilst the change  $\mathbf{w}_2 = \mathbf{e} \mathbf{x}$  represents a pure deformation. Since only three independent quantities are needed to specify the antisymmetric tensor  $\boldsymbol{\omega}$ , the elements  $\omega_{ij}$  ( $i \neq j$ ) can also be regarded as the components of an axial vector

$$\boldsymbol{\omega} = [I; \boldsymbol{\omega}] = [\omega_{32} \omega_{13} \omega_{21}]. \quad (10.8)$$

The linear transformation  $\mathbf{w}_1 = \boldsymbol{\omega} \mathbf{x}$ , where  $\boldsymbol{\omega}$  is the representation of the antisymmetric tensor, may also be written

$$\mathbf{w}_1 = \boldsymbol{\omega} \wedge \mathbf{x}, \quad (10.9)$$

where  $\boldsymbol{\omega}$  is the vector of (10.8), as may readily be seen by comparing the coefficients of  $\mathbf{w}_1$ . It is unimportant whether we use the tensor or vector methods of representing the small rotation; note that the components of the vector give the component rotations about the three axes. For the infinitesimal rotations considered in linear elasticity theory, the component rotations may be considered as vectors, and added to give the net rotation.

The geometrical interpretation of the components of the strain tensor is readily obtained. Neglecting products of these components, we find that the diagonal elements represent the extensions (i.e. changes in length per unit length) of lines originally parallel to the coordinate axes. The components  $e_{ij}$  ( $i \neq j$ ) give half the cosines of the angles between vectors originally parallel to  $x_i$  and  $x_j$ ; since  $e_{ij}$  is small,  $2e_{ij}$  thus gives the relative rotation (the change in the mutual orientation) of such vectors. The quantities  $2e_{ij}$  are commonly called the shear strains.

During the deformation, a vector of original length  $|\mathbf{x}|$  changes into a vector of length  $|\mathbf{y}| = |\mathbf{x}| + \delta |\mathbf{x}|$ . Since  $\delta |\mathbf{x}|$  is small,  $|\mathbf{y}|^2 - |\mathbf{x}|^2 = 2|\mathbf{x}| \delta |\mathbf{x}|$ , and

$$2|\mathbf{x}| \delta |\mathbf{x}| = \mathbf{x}' S' S \mathbf{x} - \mathbf{x}' \mathbf{x} = (S_{ij} + S_{ji} - 2\delta_{ij}) x_i x_j,$$

where cross products of the terms of  $S$  have again been neglected. Since  $S_{ij} + S_{ji} = 2(e_{ij} + \delta_{ij})$ , we have

$$|\mathbf{x}| \delta |\mathbf{x}| = e_{ij} x_i x_j.$$

Consider the surface

$$e_{ij} x_i x_j = \text{const.} \tag{10.10}$$

By comparison with the previous equation, we see that

$$\frac{\delta |\mathbf{x}|}{|\mathbf{x}|} = \frac{\text{const}}{|\mathbf{x}|^2},$$

where the coordinates of the end point of  $\mathbf{x}$  satisfy eqn. (10.10). The surface (10.10) thus has the property that the extension of any vector is proportional to the inverse square of the radius vector to the surface in the corresponding direction. The direction of the displacement is normal to the tangent plane to the surface at the point  $\mathbf{x}$ . We have already met this equation in the theory of finite deformations, and we noted there that  $\delta |\mathbf{x}|$  should strictly be replaced by the resolved elongation of  $\mathbf{x}$ . When the displacements are infinitesimal, the elongation quadric is known as the strain quadric, and its axes are the principal axes of the strain.

So far, we have assumed that the deformations, though infinitesimal, are homogeneous. The theory may be extended to inhomogeneous infinitesimal deformations in the following way. A point with coordinates  $x_i$  will move during deformation to a new position  $y_i$ , and the displacement may again be specified by

$$\mathbf{w} = \mathbf{y} - \mathbf{x}.$$

Consider a neighbouring point with coordinates  $x_i + \zeta_i$  (Fig. 2.8), so that before deformation the two points were related by a vector  $\zeta = [\zeta_i]$ . This point will move to a position

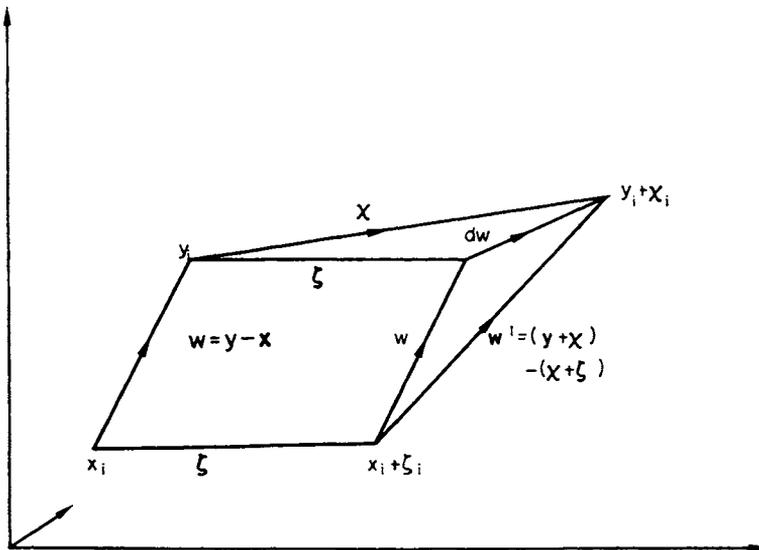


FIG. 2.8. Displacements in infinitesimal deformation.

with coordinates  $y_i + \chi_i$ , so that its displacement vector is  $\mathbf{w}' = (\mathbf{y} + \boldsymbol{\chi}) - (\mathbf{x} + \boldsymbol{\zeta})$ . We may write the components of this displacement as a Taylor series

$$(y_i + \chi_i) - (x_i + \zeta_i) = (y_i - x_i) + \frac{\partial(y_i - x_i)}{\partial x_j} \zeta_j + O(\zeta_j^2), \quad \text{etc.}$$

The differentials in this expression are to be evaluated at the point  $\mathbf{x}$ . If the components of  $\boldsymbol{\zeta}$  are small enough for the squares to be neglected, this gives

$$\chi_i - \zeta_i = \frac{\partial w_i}{\partial x_j} \zeta_j,$$

and the change in the separation of the two points is

$$d\mathbf{w} = \mathbf{w}' - \mathbf{w} = \boldsymbol{\chi} - \boldsymbol{\zeta} = \mathbf{Q} \boldsymbol{\zeta}, \tag{10.11}$$

where the matrix  $\mathbf{Q}$  has components  $\partial w_i / \partial x_j$ . Comparison of eqn. (10.11) with (10.4) shows that they are of the same form; the vector  $d\mathbf{w}$  represents the displacement of points near  $\mathbf{x}$  relative to an origin which moves with  $\mathbf{x}$ . The transformation is affine in the small region round  $\mathbf{x}$ , and  $\mathbf{Q}$  may again be separated into a strain tensor  $\mathbf{e}$  and a rotation tensor  $\boldsymbol{\omega}$ , where now

$$\left. \begin{aligned} e_{ij} &= \frac{1}{2}(\partial w_i / \partial x_j + \partial w_j / \partial x_i), \\ \omega_{ij} &= \frac{1}{2}(\partial w_i / \partial x_j - \partial w_j / \partial x_i). \end{aligned} \right\} \tag{10.12}$$

In the region around  $\mathbf{x}$ ,  $\mathbf{e}$  and  $\boldsymbol{\omega}$  may be interpreted in the same way as we previously interpreted them for the whole crystal. An inhomogeneous deformation may thus still be considered affine in a small region, and is characterized by the pure strain and rotation of each region. In contrast to the previous results, the components of  $\mathbf{e}$  and  $\boldsymbol{\omega}$  are not constants, but are functions of the coordinates  $x_i$ .

If we use the vector representation of the rotation  $\boldsymbol{\omega}$ , we have the components of this vector as

$$\omega_i = \frac{1}{2}(\partial w_k / \partial x_j - \partial w_j / \partial x_k) \quad (i \neq j \neq k)$$

or 
$$2\boldsymbol{\omega} = \text{curl } \mathbf{w}. \tag{10.13}$$

The rotation is then represented by a vector which is a function of the coordinates  $x_i$ , but is independent of the choice of coordinate system.

Utilizing the result on p. 48, it follows from (10.11) that the ratio of the new volume of a small region to its old volume is given by

$$\frac{v + \Delta v}{v} = |1 + \mathbf{Q}| = 1 + \frac{\partial w_i}{\partial x_i},$$

since all the remaining terms vanish for an infinitesimal deformation. The quantity

$$\text{div } \mathbf{w} = \frac{\partial w_i}{\partial x_i} = e_{ii} = \frac{\Delta v}{v} = \Delta, \tag{10.14}$$

is called the elastic dilatation, and may be given the same symbol  $\Delta$ , already used for finite homogeneous deformations. For infinitesimal deformations,  $\Delta$  is clearly a scalar property of the vector  $\mathbf{w}$ , and is independent of the coordinate system. If  $\mathbf{e}$  is any representation of the strain tensor  $\mathbf{e}$ , then

$$\Delta = \text{trace}(\mathbf{e}) = e_{ii}. \quad (10.15)$$

The components of  $\mathbf{e}$  are not able to vary in an arbitrary manner, since they determine the displacements  $\mathbf{w}$  of the points of the material. Provided the volume considered is singly connected (i.e. any closed curve can be shrunk continuously down to a point without crossing the boundaries),  $\mathbf{w}$  must be everywhere single-valued and continuous. This restriction results in six equations which must be satisfied by the second differential coefficients of the tensor components  $e_{ij}$ ; they are known as the equations of compatibility. We shall not derive them here, but they may all be expressed in the shortened form

$$\frac{\partial^2 e_{ij}}{\partial x_k \partial x_l} + \frac{\partial^2 e_{kl}}{\partial x_i \partial x_j} - \frac{\partial^2 e_{ik}}{\partial x_j \partial x_l} - \frac{\partial^2 e_{jl}}{\partial x_i \partial x_k} = 0. \quad (10.16)$$

## 11. STRESS-STRAIN RELATIONS: THEORY OF ELASTICITY

The equations of elastic equilibrium do not, of course, form part of the geometry of crystal lattices, but it is convenient to consider them here, since the theory of linear elasticity is so dependent on the notion of infinitesimal strain introduced in the last section. We first clarify the concept of stress. Any volume element  $v$  of a body is subject in general to forces of two kinds. There are first forces which act on all the particles of the volume element, and these are called body forces. They are caused by the presence of some external field of force, for example the gravitational field. In general, this force will not be uniform, but its effect on the whole of the volume element may be summed to give a resultant body force  $\int \mathbf{g} dv$ , where  $\mathbf{g}$  is the body force vector, and a resultant moment about any origin of  $\int (\mathbf{g} \wedge \mathbf{x}) dv$ , where  $\mathbf{x}$  is the position vector with respect to the origin.

The second type of force acting on the volume  $v$  arises from internal forces between particles of the material. Thus, consider a small planar element  $\delta O$  of the (interior) surface  $O$  separating  $v$  from the rest of the body. In general, the two parts of the body on either side of  $\delta O$  will be in a strained condition, and will exert equal and opposite forces on each other. These forces are used to specify the state of stress at a point within the element  $\delta O$ . Let the outward normal to  $\delta O$  be  $\mathbf{n}$ , a unit vector. In this section, we use  $\mathbf{n}$  rather than  $\mathbf{h}$  as the symbol for a vector normal to a plane, since we are interested in both rational and irrational planes. For the most part, we are not concerned with the crystallographic nature of the structure, so we take  $\mathbf{n}$  as a unit normal. The forces exerted by the material on one side of the element  $\delta O$  across  $\delta O$  will then reduce to a single force acting at a point within the element, and a moment about some axis. As  $\delta O \rightarrow 0$ , the direction of the resultant force approaches a fixed value, though the magnitude of both force and couple tend, of course, to zero. However, the force divided by the area remains finite, and approaches a

limit as  $\delta O \rightarrow 0$ , whilst the moment divided by  $\delta O$  tends to zero. The limiting vector, having the dimensions force per unit area, is called the stress vector  $\mathbf{p}$ , or simply the stress acting across  $\delta O$ . In a fluid, the only surface force is a uniform hydrostatic pressure, and  $\mathbf{p}$  is necessarily along the direction  $-\mathbf{n}$ . In a solid, however,  $\mathbf{p}$  may make any angle with the direction  $\mathbf{n}$ . The quantity  $\mathbf{p} \delta O$  when  $\delta O$  is small, is called the traction across  $\delta O$ . Forces of this second kind are called surface forces.

The resultant of the surface forces on the volume  $v$  is given by the surface integral  $\int \mathbf{p} dO$ . This follows since the internal surface forces across all elements  $\delta v$  within  $v$  cancel out. In the same way, there will be a resultant moment of  $\int (\mathbf{p} \wedge \mathbf{x}) dO$  on the origin, as a result of the surface forces.

The complete specification of the surface forces at a point requires a knowledge of the tractions across all planar elements at the point. We shall now show that this specification may be obtained by a set of quantities  $X_{ij}$  which form a symmetric matrix, and which are a representation of a second order tensor called the stress tensor. Consider a rectangular parallelepiped with faces perpendicular to the coordinate axes. The surface forces exerted by the material outside the parallelepiped on the material inside it are specified by a stress vector  $\mathbf{p}$  at each point of the surface. At any point on the surface normal to the axis  $x_i$ , which has a positive normal (i.e. outward normal along  $+x_i$ ), we write the stress vector as

$$\mathbf{p}_i = X_{ij} \mathbf{i}_j, \tag{11.1}$$

where  $\mathbf{i}_i$  define the orthonormal system of coordinates  $x_i$ . The quantity  $X_{(i)(i)}$  thus gives the stress acting normal to the face at the point considered, and the other two components are shear stresses.

Note that if  $\mathbf{p}$  is directed outwards from the face,  $X_{(i)(i)}$  is positive, and hence the sign convention is such that tensile stresses exerted by the surrounding material on the volume element are positive. Equally, if we had taken a point on the surface normal to  $\mathbf{i}_i$  and having negative surface normal, we should have written

$$\mathbf{p}_i = -X_{ij} \mathbf{i}_j,$$

so that tensile forces would again be specified by positive  $X_{(i)(i)}$ .

Now consider a small tetrahedron at the point  $P$  (Fig. 2.9), formed from three planes normal to the coordinate axes, and a fourth plane,  $ABC$ , which has unit normal  $\mathbf{n}$  and area  $\delta O$ . The areas of the other three faces of the tetrahedron are  $\delta O_i = n_i \delta O$ , where  $n_i$  are the components of  $\mathbf{n}$ . The outward normals to all these faces are negative, if the components of  $\mathbf{n}$  are all positive. In equilibrium, there will be no resultant force on the whole tetrahedron, so that

$$\int_v \mathbf{g} dv + \int_o \mathbf{p} dO = 0. \tag{11.2}$$

Suppose now that the volume of the tetrahedron is allowed to shrink continuously to zero. The forces proportional to the volume then vanish more rapidly than those proportional to the surface, and in the limit we need consider only the latter. The traction across

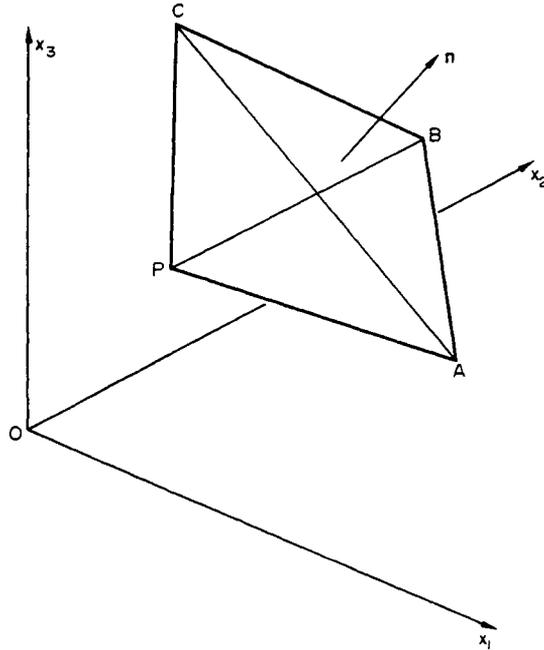


FIG. 2.9. To illustrate the meaning of the stress components at  $P$ .

each plane surface of the tetrahedron approaches the limit of the product of its area and a constant stress. Thus across the surfaces normal to the axes, the sum of the forces acting is

$$p_i \delta O_i = -n_i X_{ij} j_i \delta O.$$

If we suppose that the stress vector across  $ABC$  has the value  $\mathbf{p}$  in the limit, the equilibrium condition is

$$\mathbf{p} \delta O + p_i \delta O_i = 0$$

or

$$p_j = X_{ij} n_i. \quad (11.3)$$

In matrix form, we may write this equation as

$$(\mathbf{p}; 1) = (\mathbf{n}; 1) (\mathbf{1} \mathbf{X} \mathbf{1})$$

or

$$\mathbf{p}' = \mathbf{n}' \mathbf{X}. \quad (11.4)$$

The stress vector across any plane at a point  $P$  is thus specified by the positive normal to the plane, and the array of quantities which represent the stress tensor at this point.

We now proceed to calculate the restrictive conditions which the components of  $\mathbf{X}$  must satisfy if the body is to be in internal equilibrium. The condition of equilibrium within any volume  $v$  in the material is expressed by eqn. (11.2). If we take the component of force in the direction  $i$ ,

$$\int_v g_i dv + \int_O X_{ji} n_j dO = 0.$$

The second part of this expression may be transformed into a volume integral by means of Gauss's theorem to give

$$\int_v (g_i + \partial X_{ji} / \partial x_j) dv = 0. \quad (11.5)$$

Since the region  $v$  is arbitrary, the expression in the integral must vanish identically, and we therefore have a set of equations

$$\partial X_{ji} / \partial x_j = -g_i. \quad (11.6)$$

which are known as the equations of equilibrium.

In many important physical problems the particles of the body are not held in static equilibrium but are in states of motion. Returning to Fig. 2.9 we see that if there is a resultant force on the tetrahedron, its centre of mass will have an acceleration in the direction of this resultant. Taking the component of acceleration in the direction  $\mathbf{i}_i$ , eqn. (11.5) is replaced by

$$\int_v (g_i + \partial X_{ji} / \partial x_j) dv = \rho v \partial^2 x_i / \partial t^2,$$

where  $\rho$  is the density of the material at the point  $P$ . This gives a set of equations

$$\frac{\partial X_{ji}}{\partial x_j} + g_i = \rho \frac{\partial^2 x_i}{\partial t^2}, \quad (11.7)$$

which are known as the equations of motion.

When the body is in equilibrium, the resultant moment of the forces acting on  $v$  must vanish at any point. Choose an arbitrary origin, and let  $\mathbf{x}$  be the position vector of some point within the volume  $v$ . The moment about the origin is zero if

$$\int_v (\mathbf{g} \wedge \mathbf{x}) dv + \int_O (\mathbf{p} \wedge \mathbf{x}) dO = 0, \quad (11.8)$$

and taking the component in the direction  $\mathbf{i}_1$ ,

$$\int_v (x_3 g_2 - x_2 g_3) dv + \int_O (x_3 X_{12} n_1 - x_2 X_{13} n_1) dO = 0. \quad (11.9)$$

By using Gauss's theorem and eqn. (11.6), the second part of this expression may be written as the volume integral:

$$\int_v [-(g_2 x_3 - g_3 x_2) + X_{32} - X_{23}] dv,$$

and substituting this into (11.9), finally,

$$X_{23} = X_{32}. \quad (11.10)$$

By similarly writing down the other resolved components of (11.8), we can show that the quantities  $X_{ij}$  form a symmetric matrix with  $X_{ij} = X_{ji}$ . Equation (11.4) can thus be written in the equivalent form

$$\mathbf{p} = \mathbf{X} \mathbf{n}. \quad (11.11)$$

Using a suitable orthogonal transformation,  $\mathbf{X}$  may be referred to a new set of axes in which it has diagonal form. The values of the diagonal elements, i.e. the eigenvalues of  $\mathbf{X}$ , are called the principal stresses.

In the same way as we defined the strain quadric, we may define a surface by the relation

$$X_{ij}x_i x_j = \text{const}, \quad (11.12)$$

and this is called the stress quadric. It has the property that if a vector is drawn from the origin to a point  $\mathbf{x}$  of the stress quadric in the direction of the normal  $\mathbf{n}$  to any plane, the stress vector acting across that plane is in the direction of the normal to the tangent surface at  $\mathbf{x}$ , and the normal component of the stress vector is proportional to the inverse square of the radius vector  $\mathbf{x}$ .

For any body in equilibrium under the action of external surface forces and body forces, the six components of the stress tensor must satisfy the three eqns. (11.6) at all points in the interior. In addition, eqn. (11.11) must be satisfied over the external surface,  $\mathbf{p}$  now being the externally applied stresses. This gives three boundary conditions. These equations are insufficient to determine the state of stress; the information which is still required is the connection between the state of stress and the state of strain. This relation is obtained in the generalized form of Hooke's law, which states that the stress components are linear functions of the strain components, and vice versa. For our purpose, this is best regarded as an empirical law, based on experiment; it is approximately valid for small deformations.

Since the stress components  $X$  and the strain components  $e$  are both representations of second rank tensors, the general linear relation is of the form

$$X_{ij} = c_{ijkl}e_{kl}, \quad e_{ij} = s_{ijkl}X_{kl}, \quad (11.13)$$

where the quantities  $c_{ijkl}$ ,  $s_{ijkl}$  are representations of fourth rank tensors. Fortunately, however, the 81 components of the representations  $c$  and  $s$  are readily reduced to more manageable numbers. In the first place, both  $X$  and  $e$  are symmetric matrices with only 6 independent components. This reduces the number of independent quantities in  $c$  and  $s$  to 36. It is usual to adopt a simplified notation, due to Voigt, and write the independent components of  $X$  and  $e$  as

$$X = \begin{pmatrix} X_1 & X_4 & X_5 \\ X_4 & X_2 & X_6 \\ X_5 & X_6 & X_3 \end{pmatrix} \quad \text{and} \quad e = \begin{pmatrix} e_1 & \frac{1}{2}e_4 & \frac{1}{2}e_5 \\ \frac{1}{2}e_4 & e_2 & \frac{1}{2}e_6 \\ \frac{1}{2}e_5 & \frac{1}{2}e_6 & e_3 \end{pmatrix}. \quad (11.14)$$

The factors of one half are introduced into the off-diagonal elements of the strain tensor to conform to long-established standard notation. Before the formal theory of tensors was developed, the shear strain components  $e_4$ ,  $e_5$  and  $e_6$  were defined so as to give directly the changes in mutual inclination of vectors along the coordinate axes. These quantities are commonly called the shear strains, or simply the shears; to avoid confusion, it is better to designate them the engineering shear strains, and to distinguish them from the tensor shear strains  $e_{ij}$  ( $i \neq j$ ). Note that the array of engineering strains

$$\begin{pmatrix} e_1 & e_4 & e_5 \\ e_4 & e_2 & e_6 \\ e_5 & e_6 & e_3 \end{pmatrix}$$

does not constitute a tensor, and it is thus rather unfortunate that the older terminology is so well established. Although it is now common practice to use the tensor representation, the older component strains have been retained in defining the relations between stress and strain, which may now be written

$$\left. \begin{aligned} X_i &= c_{ij}e_j \\ e_i &= s_{ij}X_j \end{aligned} \right\} \quad (i, j = 1, 2, \dots, 6), \quad (11.15)$$

where, in contrast to previous equations, the range of  $i$  and  $j$  is from 1 to 6. The quantities  $c_{ij}$  and  $s_{ij}$  have been variously called elastic constants, moduli, or coefficients, with little agreement amongst different authors about which names are appropriate to the two sets. All these names are avoided in the modern (and descriptive) American terminology, in which the quantities  $c_{ij}$  are called stiffness constants, and the quantities  $s_{ij}$  are called compliances. In general, it is easy to examine a body under a uniaxial stress, but it is not possible to produce a uniaxial strain. The compliances may thus be determined directly by experiment; the stiffness constants only indirectly. With the above notation, the tensors  $c$  and  $s$  may be represented by  $6 \times 6$  square matrices  $c$  and  $s$ .

Now suppose the deformation at any point is changed by an infinitesimal amount, so that the strains vary from  $e_i$  to  $e_i + de_i$ . During this change, work is done by the external forces, and the potential energy in the deformed region increases. The work per unit volume may be obtained by summing the products of each stress component and the corresponding change in strain, as may be seen by considering first a unit cube of material with only one stress component,  $X_i$ , acting. If the state of strain changes by an infinitesimal amount, the only work done by  $X_i$  is  $X_i de_i$ , where  $e_i$  is the corresponding strain component. The total work done on the unit cube in increasing the stress from 0 to  $X_i$  whilst the corresponding strain component increases from 0 to  $e_i$  is thus  $\frac{1}{2}X_i e_i$ . The principle of superposition now allows us to treat the work done by a system of forces acting on a unit cube as the sum of the amounts attributable to each stress component acting separately, so that for an incremental strain, the increase in strain energy per unit volume may be written

$$dW_s/v = X_i de_i \quad (i = 1, \dots, 6). \quad (11.16)$$

The strain energy  $W_s$  must be a single valued function of the strains, so that

$$dW_s = \frac{\partial W_s}{\partial e_i} de_i$$

and the condition for  $dW_s$  to be a perfect differential is

$$\frac{\partial^2 W_s}{\partial e_i \partial e_j} = \frac{\partial^2 W_s}{\partial e_j \partial e_i},$$

i.e. from (11.16)

$$\partial X_i / \partial e_j = \partial X_j / \partial e_i.$$

From eqn. (11.15), we see that this implies

$$c_{ij} = c_{ji} \quad (11.17)$$

and the stiffness constants thus form a symmetric matrix. Similarly, the compliances  $s_{ij}$  form a symmetric matrix  $s$ . The form of the strain energy function may be written

$$W_s/v = \frac{1}{2}c_{ij}e_i e_j + \text{const.} \quad (11.18)$$

The constant is zero if the zero energy is the undeformed state. If  $W_s$  is known, the components of the stress tensor may be obtained from the relation

$$X_i = \frac{1}{v} \frac{\partial W_s}{\partial e_i} = \frac{1}{2}(c_{ij} + c_{ji})e_j = c_{ij}e_j. \quad (11.19)$$

Leaving out the constant, the quantity  $W_s/v$  gives the strain energy per unit volume stored in the material. We note in passing that we have not specified the conditions for the change  $dW_s$ . The deformation may be carried out either adiabatically or isothermally; the function  $W_s$  will be different in the two cases, and the corresponding stiffness and compliance constants are also different.

The symmetry of the components of  $c$  and  $s$  reduces the number of independent quantities in each to 21, and this number of parameters is needed to specify the elastic properties of crystals of the lowest symmetry. Further reductions in the number of independent terms are due either to the symmetry properties of the crystal structure, or to an assumed law of force. A possible assumption is that the forces between atoms are all central, i.e. they act along the lines joining two atoms, and are a function only of the separation of these atoms. If all forces are of this kind, and in addition the crystal structure is such that each atom is at a centre of symmetry, the following equations may be derived:

$$\left. \begin{aligned} c_{44} = c_{23}, \quad c_{55} = c_{31}, \quad c_{66} = c_{12}, \\ c_{56} = c_{14}, \quad c_{64} = c_{25}, \quad c_{45} = c_{36}. \end{aligned} \right\} \quad (11.20)$$

For cubic crystals, as shown below, the equations all reduce to  $c_{44} = c_{12}$ . The equations are known as the Cauchy relations; they are not valid for metals.

The existence of symmetry elements in a crystal structure leads to a reduction in the number of independent elastic stiffness constants. Thus an  $n$ -fold axis of symmetry means that the crystal is brought into self-coincidence by a rotation of  $2\pi/n$  about the axis, and such a rotation must therefore leave the elastic properties unchanged. If, therefore, we refer  $X$  and  $e$  to the new axes obtained by such a rotation, we obtain relations between the  $c_{ij}$  (or the  $s_{ij}$ ). We shall not work through the examples, but merely quote the results.

If the  $x_3$  axis is a twofold axis of symmetry (the  $x_1x_2$  plane is a plane of symmetry), then

$$c_{14} = c_{15} = c_{24} = c_{25} = c_{34} = c_{35} = c_{64} = c_{65} = 0. \quad (11.21)$$

This symmetry element thus reduces the number of independent stiffness constants to 13. If the  $x_3$  axis is a threefold axis of symmetry

$$\left. \begin{aligned} c_{16} = c_{26} = c_{34} = c_{35} = c_{36} = c_{45} = 0, \\ c_{11} = c_{22}, \quad c_{13} = c_{23}, \quad c_{14} = -c_{24} = c_{56}, \\ c_{15} = -c_{25} = -c_{46}, \quad c_{44} = c_{55}, \quad c_{66} = \frac{1}{2}(c_{11} - c_{12}), \end{aligned} \right\} \quad (11.22)$$

and there are seven independent constants. For a fourfold axis along  $x_3$ , the following relations are additional to those for a twofold axis:

$$c_{36} = c_{45} = 0, \quad c_{11} = c_{22}, \quad c_{13} = c_{23}, \quad c_{16} = -c_{26}, \quad c_{44} = c_{55}, \quad (11.23)$$

which again gives seven independent constants. A sixfold axis about  $x_3$  is obtained by combining the two- and threefold axis relations, and leaves five stiffness constants.

With the aid of these equations, the form of the matrix  $c$  for any symmetry can be determined. Thus a cubic crystal has fourfold axes  $x_1$ ,  $x_2$  and  $x_3$ . If the coordinate axes coincide with the edges of the cubic unit cell, the matrix has the form

$$c = \begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix} \quad (11.24)$$

and there are thus only three independent components. The behaviour of cubic crystals is most readily considered in relation to the combinations  $c_{44}$ ,  $(c_{11} - c_{12})/2$  and  $(c_{11} + 2c_{12})/3$ . These measure respectively the resistance to deformation when a shearing stress is applied across a  $\{100\}$  plane in a  $\langle 010 \rangle$  direction, across a  $\{110\}$  plane in a  $\langle 110 \rangle$  direction, and when a uniform hydrostatic pressure is applied to the crystal. The cubic elastic properties are thus represented by two shear moduli and a bulk modulus. The two shear moduli become equal in an elastically isotropic material, and the ratio  $2c_{44}/(c_{11} - c_{12})$  may thus be used as an elastic anisotropy factor for cubic crystals. Values of this factor are shown in Table IV, and provide an estimate of the validity of calculations using isotropic elastic theory for the metals concerned.

In an elastically isotropic body, the stiffness constants and compliances must be invariant with respect to any rotation of the axes. This leads to the further relation  $c_{44} = (c_{11} - c_{12})/2$ , and there are only two independent quantities to be specified. These are usually given the symbols

$$\mu = c_{44}, \quad \lambda = c_{12}.$$

$\mu$  is called the shear modulus;  $\lambda$  has no special name. If we now write out the components of stress for the isotropic medium, we find they can be expressed in the form

$$\begin{aligned} X_i &= c_{ij}e_j = \lambda(e_1 + e_2 + e_3) + 2\mu e_i = \lambda\Delta + 2\mu e_i & (i = 1, 2, 3) \\ \text{and} \quad X_i &= c_{(i)(i)}e_i = \mu e_i & (i = 4, 5, 6). \end{aligned} \quad (11.25)$$

Using eqn. (11.14) to transform back to the components of the stress and strain tensors,

$$X_{ij} = \lambda\delta_{ij}\Delta + 2\mu e_{ij}, \quad (11.26)$$

and, in particular,

$$X_{ii} = X_{11} + X_{22} + X_{33} = (3\lambda + 2\mu)\Delta.$$

TABLE IV. VALUES OF ANISOTROPY FACTOR,  $2c_{44}/(c_{11}-c_{12})$  FOR CUBIC METALS

(Elastic stiffnesses taken from Huntington, 1958)

Metal	Structure	Anisotropy factor
Lithium	b.c.c.	9.39
Sodium	b.c.c.	8.14
Potassium	b.c.c.	6.34
Copper	f.c.c.	3.20
Silver	f.c.c.	3.01
Gold	f.c.c.	2.90
Aluminium	f.c.c.	1.29
Silicon	Diamond cubic	1.56
Germanium	Diamond cubic	1.66
Lead	f.c.c.	3.89
Nickel	f.c.c.	2.51
Thorium	f.c.c.	3.62
Molybdenum	b.c.c.	0.77
Tungsten	b.c.c.	1.00
$\beta$ -Brass	b.c.c.	8.40

The components of the strain tensor,  $e_{ij}$ , may similarly be obtained in terms of the stress components, using (11.13) which gives

$$e_{ij} = -\frac{\lambda}{2\mu(3\lambda+2\mu)} \delta_{ij} X_{kk} + \frac{1}{2\mu} X_{ij}. \quad (11.27)$$

If  $e$  is a diagonal matrix,  $X$  is also diagonal, so that the principal axes of stress and strain must coincide for an isotropic body.

Real crystals are not elastically isotropic, although some cubic crystals have small anisotropy factors. However, very many important problems are concerned not with the properties of isolated crystals, but with polycrystalline aggregates. If the distribution of orientations is effectively random, and the grain size small enough, these assemblies behave as isotropic bodies. Experimental results on elastic properties are then usually expressed in the form of various moduli of elasticity. If the body is deformed by a uniaxial stress, then all the  $X_{ij}$  except  $X_{11}$  are zero. The corresponding strains are

$$e_{11} = \frac{\lambda + \mu}{\mu(3\lambda + 2\mu)} X_{11}, \quad e_{22} = e_{33} = -\frac{\lambda}{2\mu(3\lambda + 2\mu)} X_{11}, \quad e_{ij} = 0 \quad (i \neq j).$$

The ratio  $Y = X_{11}/e_{11} = \mu(3\lambda + 2\mu)/(\lambda + \mu)$  is called Young's modulus of elasticity. The ratio of the lateral contraction to the longitudinal expansion in such an experiment,  $\nu = -e_{22}/e_{11} = \lambda/2(\lambda + \mu)$  is called Poisson's ratio.

A body subject to pure shear has  $X_{23} = X_{32}$ , and all the other components of stress are zero. This gives

$$e_{23} = e_{32} = X_{23}/2\mu.$$

The quantity  $\mu$  thus represents the ratio of the shearing stress to the change in angle between the  $x_2$  and  $x_3$  axes; this latter quantity is the (engineering) shear strain, and  $\mu$  is the shear

modulus. It is also readily seen that when a uniform hydrostatic pressure is applied, so that  $X_{11} = X_{22} = X_{33}$ , the ratio of the compressive stress to the cubical compression,  $-\Delta$ , is given by  $K' = \lambda + 2\mu/3$ , where  $K$  is called the bulk modulus.

The theory of elasticity has to be applied to two principal classes of problem. These are the calculation of the distribution of stress and the displacements in the interior of an elastic body when the body forces are known, and either (a) the displacements over the surface of the body or (b) the forces acting on the surface, are specified. The equations to be solved are the stress equations of equilibrium (in more general problems, the equations of motion), together with the auxiliary conditions represented by the stress-strain relations, and where required (in the second type of problem), the compatibility conditions. For isotropic media, we may substitute eqns. (11.26) into the equations of equilibrium, to give

$$(\lambda + \mu) \frac{\partial \Delta}{\partial x_i} + \mu \nabla^2 w_i = -g_i \tag{11.28}$$

or, using (10.14),

$$(\lambda + \mu) \text{grad div } \mathbf{w} + \mu \nabla^2 \mathbf{w} = -\mathbf{g}. \tag{11.29}$$

Equations (11.28) and (11.29) are very useful in many elementary problems of the first type. Sometimes, however, it is convenient to use the identity  $\nabla^2 = \text{grad div} - \text{curl curl}$  to transform (11.29) into

$$\left. \begin{aligned} (\lambda + 2\mu) \text{grad div } \mathbf{w} - \mu \text{curl curl } \mathbf{w} &= -\mathbf{g} \\ (\lambda + 2\mu) \text{grad } \Delta - \mu \text{curl } \boldsymbol{\omega} &= -\mathbf{g}. \end{aligned} \right\} \tag{11.30}$$

The first problem is completely solved if solutions of one of the equivalent forms (11.28)–(11.30) are obtained, subject to the boundary conditions in which  $\mathbf{w}$  is specified over the limiting surface of the solid considered.

To solve the second type of problem, the differential equations have to be expressed in terms of stresses rather than displacements. Since not every solution of the stress equations of equilibrium represents a possible state of strain, it is necessary to incorporate the compatibility equations. After some manipulation, eqns. (10.16) and (11.27) give the following set of six differential equations:

$$\nabla^2 X_{ii} + \frac{1}{1+\nu} \frac{\partial^2}{\partial x_i \partial x_j} (X_{ii}) = -\frac{\nu}{1+\nu} \delta_{ij} \text{div } \mathbf{g} - \left( \frac{\partial g_i}{\partial x_j} + \frac{\partial g_j}{\partial x_i} \right). \tag{11.31}$$

These are known as the Beltrami–Michell compatibility equations. The second main class of elastic problem then involves the solution of these equations, subject to the boundary conditions in which the forces acting over the surface of the body are specified. In this book, we shall not be concerned with problems of this type, nor with the mixed class when the boundary conditions are specified partially as forces and partially as displacements.

## REFERENCES

- BOWLES, J. S. and MACKENZIE, J. K. (1954) *Acta metall.* **2**, 129, 138.  
BULLOUGH, R. and BILBY, B. A. (1956) *Proc. Phys. Soc. B*, **69**, 1276.  
CAHN, R. W. (1953) *Acta metall.* **1**, 49.  
CROCKER, A. G. (1962) *Phil. Mag.* **7**, 1901.  
FRANK, F. C. (1965) *Acta crystallogr.* **18**, 862.  
FRANK, F. C. (1988) *Met. Trans.* **19A**, 403.  
HANDSCOMB, D. C. (1958) *Can. J. Math.* **10**, 85.  
HUNTINGTON, H. B. (1958) *Solid State Phys.* **7**, 214.  
JASWON, M. A. and WHEELER, J. A. (1948) *Acta crystallogr.* **1**, 216.  
NEUMANN, P. (1966) *Phys. stat. sol.* **17**, K71.  
NICHOLAS, J. F. (1966) *Acta crystallogr.* **21**, 880; (1970); *Phys. stat. sol.* (a) **1**, 563.  
NICHOLAS, J. F. and SEGALL, R. L. (1970) *Acta crystallogr.* **A26**, 522.  
NYE, J. F. (1957) *Physical Properties of Crystals*, Clarendon Press, Oxford.  
OKAMATO, P. R. and THOMAS, G. (1968) *Phys. stat. sol.* **25**, 81.  
OTTE, H. M. and CROCKER, A. G. (1965) *Phys. stat. sol.* **9**, 44; (1966) *Ibid.* **16**, K25.  
SOKOLNIKOFF, I. S. (1946) *Mathematical Theory of Elasticity*, McGraw-Hill, New York; (1951) *Tensor Analysis*, Wiley, New York.  
ZACHARIASEN, W. H. (1945) *Theory of X-Ray Diffraction in Crystals*, Wiley, New York.

## CHAPTER 3

# *The Theory of Reaction Rates*

### 12. CHEMICAL KINETICS AND ACTIVATION ENERGY

A typical problem in the theory of transformation is the motion of an interface separating two regions of different composition. Provided we interpret the term "chemical reaction" in a rather more liberal sense than is usual, we may regard the growth of the more stable region as the combination of a transport process (atomic diffusion) with a chemical reaction (formation of a new phase). Moreover, the transport process may itself be regarded as a series of chemical reactions, and this concept has proved very useful. Under appropriate circumstances, we may use theories which have been developed to describe the kinetics of chemical reactions and extend them to analogous phenomena such as diffusion.

Most descriptions of chemical kinetics are based on a statistical theory developed by Eyring and his co-workers (Glasstone *et al.*, 1941). This theory depends on the assignation of thermodynamic properties to non-equilibrium states, and is now usually known as the "absolute reaction rate theory" or as the "transition state theory". Despite its success in chemical problems, the claim implicit in the first name seems a little too ambitious, since calculation of reaction rates from first principles is seldom possible in practice. When applied to metallic assemblies, the method has yielded useful results, but we must emphasize at once that there are many solid state phenomena for which it is clearly not valid. Reaction rate theory should therefore be applied only after careful consideration of the probable physical mechanism of the process considered. When this mechanism appears to be consistent with the postulates of the theory, as discussed below, the chemical analogy may be useful, but in other cases it may be misleading.

The basic assumption of Eyring's theory is that the whole reaction consists of the repetition of a fundamental unit step, or series of such steps. Each unit step arises from the interaction of a *small* number of atoms or molecules to form a new configuration; before this is achieved, the group passes through intermediate situations which have higher energies than either the initial or final states. There is a critical configuration which must be attained if the final state is to be reached, and the increase in free energy required to form this critical configuration determines the time taken to make the unit step. There is thus an activation energy for the process which must be supplied by a local fluctuation in the thermal energy of the assembly.

These conditions are obviously satisfied by most ordinary chemical reactions, both in

the gaseous phase and in liquid solution. There are many additional difficulties in the solid state, but the general description of Chapter 1 suggests that nucleation and growth transformations can usually be split up into unit steps governed by thermally supplied activation energies, at least as a first approximation. On the other hand, changes in which the movements of large numbers of atoms are closely co-ordinated are unlikely to be controlled by thermal fluctuations, so that this kind of theory is inappropriate to martensitic growth. Early attempts were occasionally made to apply the theory to phenomena to which it is unsuited and were incorrect mainly because of insufficient attention to the physical mechanisms. Thus a phenomenological model of plastic deformation in which individual units of flow are treated as single thermally activated processes is not generally valid, although there are many dislocation interactions contributing to plastic deformation in which the concept of activation energy is useful.

Once this danger has been recognized it is not difficult to select processes which are genuinely thermally activated, and to which the chemical analogy may be supposed to apply. It is very much more difficult to decide whether the rather restrictive assumptions of Eyring's theory are valid in any particular thermally activated process. Here we encounter difficulties of several kinds. The original form of the reaction rate theory applied to gases and to liquid solutions which behave ideally. Attempts to extend the theory to non-ideal assemblies are generally made by introducing activity coefficients, as in equilibrium thermodynamics, but the equations then contain quantities of rather uncertain physical significance. This problem arises in most applications of the theory to the solid state, and the logical difficulties involved are frequently ignored. Fortunately the difficulty can often be avoided by making use of a treatment due to Wert and Zener (1949) and Zener (1952), which permits a direct correlation of theory and experiment.

Another difficulty which arises with solids, and which was first emphasized by Crussard (1948), is that the atoms or molecules cannot be legitimately treated as a set of independent quantum systems, especially in problems in which thermal energy plays an important role. The simplest form of the reaction rate theory effectively considers the thermal energy of motion of a solid to be the energy of a number of independent atomic or molecular oscillators, whereas in fact the motions are coupled so that much of the energy may be contained in relatively long wavelength vibrations. Thus in considering the probability of a small number of atoms or molecules surmounting an energy barrier, we should treat the energy increase as a temporary local concentration of phonons, caused by interference of wave packets. Some progress has been made in developing a theory of this kind, but most existing treatments of molecular processes in solids use the ordinary Eyring theory. This approximation should not be too seriously wrong, except at very low temperatures; it is roughly equivalent to using an Einstein model of a solid instead of a Debye model, or a more sophisticated model. Vineyard (1957) has clarified the many-body aspects of thermally activated rate processes in solids and has shown that in a harmonic approximation the formalism of the Eyring theory is still acceptable. This paper is very important in the development of a satisfactory conceptual theory, but discussion of it is deferred to the end of this chapter so that it may more readily be related to the work of Wert and Zener.

These other assumptions, which are listed in the next section, provide the final and

greatest difficulty in the application of the theory. The problem concerns the validity of thermodynamic concepts applied to the critical transition state at the top of the free energy barrier in the unit step and of the assumption that reactants in this transition state are in equilibrium with the reactants in the initial state. This requires that the lifetime of the transition state should be long in comparison with the time of thermal relaxation in the lattice, and it is doubtful if this is so. Clearly theories of the process which are based on the dynamic properties of the molecules participating in the unit step would be preferable to the pseudo-thermodynamic theories with which we usually have to be satisfied. Some progress in developing such theories has been made by Rice (1958) for the case of diffusion, and is discussed on p. 94. We may remark parenthetically that although we have kept the discussion in this section general, almost all advances in the theory of rate processes in solids have been made with reference to the theory of diffusion in the first instance.

In the remainder of this section we shall summarize some well-known results of chemical kinetics as a prelude to consideration of the rate theories in the next section. The unit step in a chemical reaction may be represented by the equation



where the reacting systems (atoms, molecules, etc.) are represented by  $A, B, \dots$ , the products by  $P, Q, \dots$ , and the numbers  $a, b, \dots, p, q, \dots$ , are integers. The rate of reaction is specified by the velocity constant or specific rate constant  $k_f$  which is so defined that the number of product systems formed per unit volume in unit time is equal to  $k_f$  multiplied by some function of the concentrations of the reacting systems. This may seem rather a vague definition; it arises because the form of the dependence of the rate of reaction on the concentrations of  $A, B$ , etc., in the general case can only be determined by experiment. However, the function is usually a simple one, and it is then possible to assign an *order* to the reaction.

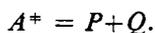
The order is defined as the power of the concentration terms which determine the rate law. Suppose, for example, that we have a single species of reacting molecule,  $A$ . If the rate of reaction is proportional to the concentration  $c_A$  (e.g. a radioactive decomposition) it is first order; if it is proportional to  $c_A^2$ , it is second order. Similarly, a gaseous reaction between two molecules  $A$  and  $B$ , is of the second order if it is proportional to  $c_A c_B$ . It follows from this that although the rate constant  $k_f$  is always numerically equal to the rate of reaction at unit concentration of each of the reactants, the physical dimensions of  $k_f$  depend on the order of the reaction.

In complex reactions involving several unit steps, it is not always possible to give an unambiguous meaning to the order of the reaction. It is then preferable to specify the order with respect to a particular reactant, this being the power of the concentration term of that reactant which enters into the rate law. In the example above, the reaction between  $A$  and  $B$  is of first order with respect to each reactant.

The molecularity of a reaction is defined as the number of reacting molecules (or atoms) which take a direct part in the change. Before reaction can occur, these molecules must be raised to higher energy states, and a particular configuration which is critical for the change is said to form the activated complex. The molecularity of the reaction can thus be stated

only when the mechanism is known, and only molecules which form part of the activated complex are considered to take a direct part in the reaction. If there are several unit steps, the molecularity can only be specified for each of these separately.

The meaning of direct in the above definition may be clarified by the following example in which we suppose the unit step to consist of the decomposition of a single molecule  $A$  into two molecules  $P$  and  $Q$ . The decomposition can occur only if the  $A$  molecule acquires a higher energy than it possesses in the ground state, and in this condition which we write  $A^*$  the molecule constitutes the activated complex. Now suppose the reaction occurs in the gaseous phase. Interchanges of energy among the molecules then result only from collisions, and the whole unit step may be written



The molecularity of this reaction is one because one of the molecules in the collision plays only an indirect role in the reaction.

If the whole reaction consists of one unit step, we might expect the molecularity to equal the order of the reaction. This is generally true, but there is one important exception which we shall mention below. When there are several atomic processes, there is usually one which is much slower than the others, and the overall reaction rate is then determined almost entirely by the rate applicable to this unit step. In this case, the overall order of the reaction will be equal to the molecularity of this rate determining step. In complex reactions there may be two or more slow steps, and the whole reaction does not then have a well-defined order, though each unit step still has a definite molecularity. The one exception to these simple rules is the monomolecular gas reaction discussed above. The kinetic theory of gases shows that this is first order at high pressures and second order at low pressures.

The distinction between reaction order which refers to the overall change, and molecularity which refers to the mechanism should be remembered, but will not concern us further. In metallic transformations we shall encounter only first-order unimolecular reactions and second-order bimolecular reactions. In the equations of this chapter we shall use an undetermined number of concentration terms in the expressions for the reaction rate in order to facilitate comparison with standard chemical textbooks. The actual number of terms must be determined for the unit step in any process considered.

A chemical reaction such as that represented by eqn. (12.1) does not usually proceed to completion, but is reversible. The assembly then tends to an equilibrium state in which the concentrations of products and reactants remain constant. This equilibrium is the result of a dynamic balance between the forward and reverse reactions, which continue on an atomic scale, but with equal velocities. There must then be a connection between the specific rate constants for the two reactions and the thermodynamic equilibrium constant which describes the macroscopic state of the assembly.

The rates of the forward and back reactions may be written  $k_f f_f(c_A, c_B, \dots)$  and  $k_b f_b(c_P, c_Q, \dots)$ , where the form of the functions  $f$  is not specified. Equating the two rates gives

$$k_f/k_b = f_f(c_A, c_B, \dots)/f_b(c_P, c_Q, \dots) = f(c_A, c_B, \dots, c_P, c_Q, \dots).$$

Under specified constraints,  $k_f$  and  $k_b$  are both constants, and it follows therefore that there must be some function of the concentrations of reactants and products which is constant at equilibrium. However, the equilibrium constant  $K$  specifies the equilibrium state in terms of the activities and hence of the concentrations of the products and reactants. It follows that  $K$  must equal  $f(c_A, c_B, \dots, c_P, c_Q, \dots)$ , or else be some power of this function. In practice  $k_f/k_b$  almost always equals  $K$ , though, in a few cases, this quotient has the value  $K^{1/2}$  or  $K^2$ .

We have assumed that there is only one reaction path leading from the initial to the final state. If there are several alternative paths, the functions  $f_f, f_b$  will contain as many separate terms as there are paths. At equilibrium, the sum of the rates from reactants to products by all paths must equal the sum of the reverse rates, i.e.

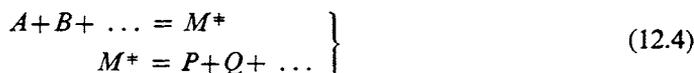
$$k_f f_f + k'_f f'_f + k''_f f''_f + \dots = k_b f_b + k'_b f'_b + k''_b f''_b + \dots \quad (12.2)$$

This equation satisfies the condition for stationary values of the thermodynamic parameters, and if the macroscopic thermodynamic laws are regarded as postulates, they do not impose any further restrictions. However, these laws may be related to the kinetic behaviour of the systems of the assembly, and statistical considerations show that (12.2) is insufficient to ensure equilibrium; in addition it is necessary that the forward rate along any one reaction path shall equal the reverse rate along that path. This condition, expressed in the equations

$$k_f/k_b = k'_f/k'_b = k''_f/k''_b = K, \quad (12.3)$$

is known as the principle of detailed balancing. It is one aspect of a more fundamental statistical law called the principle of microscopical reversibility, which we shall discuss in the next chapter. The fundamental reason for the principle of detailed balancing may be illustrated by the following argument. Let there be two reaction paths from the initial to the final condition, one of which depends on the presence of a surface catalyst, and assume that eqns. (12.3) are not satisfied at equilibrium. We may now consider a virtual change in which the catalyst is removed thereby closing one reaction path and displacing the equilibrium. This will result in a spontaneous change of the assembly to a new condition. If the catalyst is now re-introduced, there will be a further spontaneous change, and by performing such cycles repeatedly we would be able to draw energy continuously from the assembly, in contradiction to the second law of thermodynamics.

Consider now the unit step in any reaction, which we may represent symbolically as



Whatever the nature of this step, it is found experimentally that for a very large number of reactions, the rate of reaction may be expressed in the form

$$\text{i.e.} \quad \left. \begin{aligned} \ln k_f &= \ln C_4 - E/RT, \\ k_f &= C_4 \exp(-E/RT) = C_4 \exp(-\varepsilon/kT), \end{aligned} \right\} \quad (12.5)$$

where  $C_4$  is a constant and  $E$ ,  $\varepsilon$  are called the experimental activation energies per mole and per molecule respectively. This is the Arrhenius equation. Provided the temperature range considered is not large,  $C_4$  and  $E$  are effectively constants.

Consider now the energy diagram of Fig. 3.1 which shows the variation of internal or potential energy during the course of the reaction. The initial and final states are stable and hence correspond to minima in the energy field; at some intermediate stage the energy

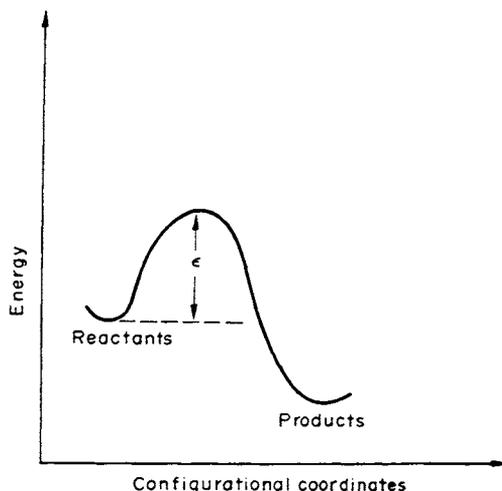


FIG. 3.1. Schematic energy diagram.

risers to a maximum. The abscissae of the figure represent configurational coordinates along the reaction path. In the simplest example the reaction might consist of the approach of two molecules along a linear path until at a critical separation they form the complex  $M^\ddagger$ . The first part of the reaction path would then be a linear spatial coordinate, with the intermolecular distance decreasing; similarly, the second part of the reaction path might be another linear spatial coordinate with increasing intermolecular distance. In more complex cases, the behaviour of the molecules cannot be visualized so readily, but there will always be some continuous series of configurations in terms of which we may interpret the abscissae of the figure.

From Fig. 3.1, we see that before the reactants can travel along the reaction path and thus form the products, they must acquire temporarily an amount of energy equal to the height of the energy barrier. If the probability of the systems having this energy were proportional to  $\exp(-\varepsilon/kT)$ , the Arrhenius equation would follow naturally. In fact this is not true, but it gives an approximate interpretation of the exponential factor of eqn. (12.5).

In order that the product  $C_4 \exp(-E/RT)$  shall have the right dimensions,  $C_4$  must have the dimensions of a frequency. The early theory of gaseous reactions assumed that  $C_4$  was the frequency of collision of the reacting molecules, collisions being necessary for transfers of energy to be possible. In the same way, for a bound particle in the solid state,  $C_4$  would be the frequency of vibration in the direction leading to the reaction path. In effect,  $C_4$  measures the number of times per second that the reacting systems can attempt to traverse

the reaction path, and the exponential term gives the probability of their having sufficient energy to be successful at any one attempt. In the next section we shall obtain more correct interpretations of both these two factors.

### 13. THE REACTION RATE THEORY

We begin this section with a description of the statistical theory of reaction rates developed by Eyring *et al.*, and take up later the special difficulties mentioned on pp. 80-1. The basic assumptions of Eyring's theory are:

(1) The reaction is characterized by some initial configuration which proceeds to the final configuration by a continuous change of coordinates. This change constitutes the unit step of the process. The initial and final configurations are stable states situated at relative minima in the energy field, so that a system in one of these states tends to return to its original position when given a small displacement.

(2) There is an energy barrier between the initial and final states along any reaction path, and the most-favoured reaction path will generally be that for which this energy barrier is least. The energy maximum along the most-favoured reaction path will be at a lower energy level than adjacent points lying on other reaction paths, and hence this maximum will be a saddle point in the energy field. It is usually tacitly assumed that there is only one such saddle point along the chosen reaction path; this statement seems reasonable, although no firm proof can be given. The energy field concerned was originally considered to be the potential energy field, but later developments have shown that we must consider free energies (see below).

(3) There is a critical configuration intermediate between the initial and final states and situated at the energy maximum (strictly the free energy maximum) of the reaction path. This configuration is called the activated complex or activated state, and once it is attained there is a high probability of the reaction proceeding to completion.

(4) The reacting systems in the initial state are regarded as being in equilibrium with the activated complexes  $M^*$ , even though these do not form equilibrium states.

(5) The activated complex is assumed to possess all the properties of a normal equilibrium configuration, except that it has no degree of freedom corresponding to vibrational motion along the reaction path. Such vibrational motion leads in fact to decomposition of the complex to form the final stable configuration. In place of this vibrational degree of freedom, we may thus suppose the complex to have an additional degree of freedom corresponding to translational motion along the reaction path.<sup>†</sup> The velocity of the reaction is determined by the rate at which products form from  $M^*$ , i.e. by the rate at which the activated complexes travel along the reaction path on the top of the energy barrier.

In order to calculate absolute reaction rates from first principles it is necessary to evaluate the energy along the reaction path. This is a very difficult problem, which has been

<sup>†</sup> This must not be confused with a spatial translational degree of freedom of the complex. In the rather strange terminology of this theory, translational motion of the complex along the reaction path is synonymous with disintegration of the complex.

approximately solved only for the very simplest reactions. We emphasize again that either experiment or other considerations must decide whether or not the above assumptions are valid for any particular reaction process. As already stated, wrong results have occasionally been obtained by application of the theory to phenomena which do not satisfy the first two assumptions, but in most cases it is relatively easy to decide whether or not they are approximately satisfied. When (1) and (2) are valid, it has generally been assumed that the remaining conditions are not unduly restrictive, the main alternative being that the activated complexes decompose as rapidly as the reacting systems accumulate the necessary thermal energy, so that there is no equilibrium.

Let us consider the average "velocity" of an activated complex in the  $+x$  direction, which we regard as the direction of the reaction path, not as a fixed spatial coordinate. A complex of effective mass  $m$  and velocity  $\dot{x}$  has kinetic energy  $\frac{1}{2}m\dot{x}^2$ . If we assume an equilibrium distribution of velocities, the probability of the complex having a velocity between  $\dot{x}$  and  $\dot{x} + d\dot{x}$  is proportional to  $\exp(-m\dot{x}^2/2kT) d\dot{x}$ , and the average velocity of a complex in the  $+x$  direction will be given by

$$\bar{\dot{x}} = \frac{\int_0^{\infty} \exp(-m\dot{x}^2/2kT) \dot{x} d\dot{x}}{\int_{-\infty}^{\infty} \exp(-m\dot{x}^2/2kT) d\dot{x}}$$

The integration limits of the dividend are 0 to  $\infty$  since we are not considering velocities in the  $-x$  direction. Evaluation of the integrals gives

$$\bar{\dot{x}} = (kT/2\pi m)^{1/2}.$$

The activated complexes may be regarded as being stable within an (unspecified) length  $\delta x$  lying along the top of the reaction path. The average time taken for a complex to cross the barrier is thus given by  $\tau = \delta x/\bar{\dot{x}}$ , and if  $c^*$  is the number of complexes per unit volume which are within the length  $\delta x$  of the reaction path, the number of complexes decomposing per unit volume per unit time is  $c^*/\tau$ . This is the rate of the unit step reaction, and is equal to the velocity constant multiplied by the concentrations of the reacting systems, i.e.

$$k_f c_A c_B \dots = (c^*/\delta x) (kT/2\pi m)^{1/2}.$$

For both ideal and non-ideal assemblies,

$$c^*/c_A c_B \dots = Q^{*||}/Q_A Q_B \dots$$

in which the quantities on the right-hand side are the partition functions for the activated complex and for the reactants, all referred to the same energy zero. If we refer each partition function to its own zero energy level, we obtain

$$c^*/c_A c_B \dots = Q^{*||} \exp(-\varepsilon_0/kT)/Q_A Q_B \dots,$$

where  $\varepsilon_0$  is the difference in the zero energy levels of the activated complex and the reacting systems, and is therefore the activation energy per system for the reaction at 0 K. We have

written the partition function for the activated complex in the form  $Q^{\ddagger}$  purely for convenience, since we wish to reserve the symbol  $Q^{\ddagger}$  for another related quantity.

We now have

$$k_f = \frac{Q^{\ddagger}}{Q_A Q_B \dots} \left( \frac{kT}{2\pi m} \right)^{1/2} \frac{\exp(-\varepsilon_0/kT)}{\delta x}. \quad (13.1)$$

$Q^{\ddagger}$  is the complete partition function for the activated complex, and may be factorized into the partition function for translational motion along the reaction path, and the partition function for the remaining degrees of freedom. We thus have

$$Q^{\ddagger} = Q^{\ddagger} Q_x^{\ddagger},$$

where if we use the standard expression for a translational partition function

$$Q_x^{\ddagger} = (2\pi mkT/h)^{1/2} \delta x. \quad (13.2)$$

Combining eqns. (13.1) and (13.2),

$$k_f = (kT/h) (Q^{\ddagger}/Q_A Q_B \dots) \exp(-\varepsilon_0/kT). \quad (13.3)$$

We note that the  $(kT/h)$  factor in this equation arises from the combination of the mean velocity in the "direction" of decomposition with the translational partition function in this direction. Since  $\delta x$  does not appear in the result, we may choose it, if we wish, so as to make  $Q_x^{\ddagger}$  equal to unity.  $Q^{\ddagger}$  and  $Q^{\ddagger}$  are then numerically equal, but they refer of course to different degrees of freedom. In calculation of  $Q^{\ddagger}$ , motion of the complex along the reaction path must be excluded. The form of the derivation shows that  $kT/h$  is a universal factor entering into all such rate equations. An alternative way of obtaining eqn. (13.3) illustrates this point. Instead of regarding the activated complex as possessing no vibrational degree of freedom along the reaction path, we may suppose that its vibration in this direction is so weak that each vibration is transformed into a translation and leads to decomposition. The rate of reaction is then equal to the frequency of vibration along the reaction path, multiplied by the concentration of activated complexes. Hence

$$\text{or} \quad \left. \begin{aligned} k_f c_A c_B \dots &= c^{\ddagger} \nu \\ k_f &= (Q^{\ddagger} / Q_A Q_B \dots) \nu \exp(-\varepsilon_0/kT). \end{aligned} \right\} \quad (13.4)$$

We now factorize  $Q^{\ddagger}$  into  $Q^{\ddagger}$  and the partition function corresponding to the vibrational freedom along the reaction path. Since the activated state is thoroughly excited, we may take the classical value of  $kT/h\nu$  for this, and combining with eqn. (13.4) we again obtain eqn. (13.3).

Eqn. (13.3) may be regarded as the fundamental equation of the reaction rate theory, and applies to both ideal and non-ideal assemblies. There is, however, a factor which we have neglected. Quantum mechanics show that there is a finite probability that a system with sufficient energy to surmount the energy barrier and which is travelling along the forward reaction path will nevertheless be reflected back to the initial state. For completeness, eqn. (13.3) should therefore be multiplied by a quantum mechanical transmission

probability. In normal reactions, this factor is close to unity and we shall ignore it. In the same way, there is a finite probability of a set of systems reaching the final state even although they do not possess the requisite activation energy. This "tunnel effect" is quite negligible except for nuclear phenomena and for extremely low temperatures.

For an assembly which behaves ideally, we may introduce an equilibrium constant  $K_c$ , where

$$K_c = (c^*/c_A c_B c_C \dots) = (Q^*/Q_A Q_B \dots) \exp(-\varepsilon_0/kT).$$

We may also define a similar "equilibrium constant"  $K_c^*$  by the equation

$$K_c^* = (Q^*/Q_A Q_B \dots) \exp(-\varepsilon_0/kT), \quad (13.5)$$

where  $K_c^*$  differs only slightly from  $K_c$ , and may indeed be made numerically equal to  $K_c$  by suitable choice of  $\delta x$ . This fact is frequently allowed to obscure the difficulty of giving a physical interpretation to  $K_c^*$ . Since the number of degrees of freedom in the divisor and dividend of eqn. (13.5) are different,  $K_c^*$  cannot be a true equilibrium constant and the definition is made purely for convenience. The constant refers to the hypothetical equilibrium between the reactants and the activated complexes which have, as it were, been deprived of a degree of freedom.

We can now write eqn. (13.3) in the form

$$k_f = (kT/h)K_c^* \quad (\text{ideal assemblies only}). \quad (13.6)$$

In fact, since we shall be more interested in applying the rate theory to solutions which are not dilute, it is often more convenient to use the equilibrium constant  $K^*$  expressed in terms of atom fractions rather than concentrations. Since the activated complex is to be regarded as a single molecule, it follows that

$$K^* = c^{i-1} K_c^*,$$

where  $i$  is the molecularity of the unit step, and  $c$  is the total concentration. This gives

$$k_f = (kT/h)c^{1-i} K^* \quad (\text{ideal assemblies only}). \quad (13.7)$$

For reactions in non-ideal solutions, the equilibrium constant is normally given in terms of activities rather than atomic fractions. Thus for the equilibrium between reactants and activated complexes,

$$k_f = (kT/h)(\gamma_A \gamma_B \dots / \gamma^*) c^{1-i} K^*. \quad (13.8)$$

where the  $\gamma$  are activity coefficients. This equation, though quoted in standard textbooks and papers, raises the problem of the meaning to be attributed to  $\gamma^*$ , and this difficulty may be linked with that of defining  $K^*$ . A better formalism which provides a more direct link with experimental observations was introduced by Wert and Zener (1949) in connection with the theory of diffusion; we consider this below.

By means of the thermodynamic equation

$$RT \ln K^* = -\Delta G^*$$

we can express the reaction rate in the form

$$\begin{aligned} k_f &= (kT/h) (\gamma_A \gamma_B \dots / c^{i-1} \gamma^*) \exp(-\Delta G^*/RT) \\ &= (kT/h) (\gamma_A \gamma_B \dots / c^{i-1} \gamma^*) \exp(\Delta S^*/R) \exp(-\Delta H^*/RT). \end{aligned} \quad (13.9)$$

The quantities  $\Delta G^*$ ,  $\Delta S^*$ ,  $\Delta H^*$  are usually called the molar free energy, molar entropy and molar heat of activation respectively. They are then stated to be the differences in free energies, entropies, and enthalpies of the activated complexes and the reactants when all are in their standard states.<sup>†</sup> It follows from the discussion above, however, that their physical meaning is rather less certain than this statement suggests. The standard state of the activated complexes is not well defined, and must refer to complexes with one degree of freedom less than the reactants. Nevertheless, this equation is perhaps the most useful form of expressing the result of the rate theory, and the physical significance is certainly clearer than in the expression involving partition functions. We note especially that the important rate determining factor is the free energy of activation.

In some particular examples it is possible to obtain an expression with a more precisely defined physical meaning. If we return to eqn. (13.3), we may factorize the divisor on the right-hand side by taking out the partition function which corresponds to vibrational motion of the reactants along the reaction path. The quotient of the two kinds of partition function remaining may then be put equal to  $\exp(-\Delta_a G/RT)$ , where  $\Delta_a G$  differs from  $\Delta G^*$  defined above. We then obtain

$$k_f = (kT/h) (1/Q_v) \exp(-\Delta_a G/RT), \quad (13.10)$$

where  $Q_v$  is the vibrational partition function of the reactants.  $\Delta_a G$  has a physical interpretation. It is the isothermal work needed slowly to transfer the configuration of molecules which form the reactants from their initial positions to their critical positions, subject to the restriction that they are allowed at all times to vibrate only in the plane perpendicular to the reaction path. Moreover, the effect of the activity coefficients is then included in the definition of  $\Delta_a G$ , so that eqn. (13.10) is valid for both ideal and non-ideal solutions.

Equation (13.10) was obtained for the particular case of diffusion by Wert and Zener, as already mentioned. Its utility obviously depends on whether a simple expression is available for  $Q_v$ , and in chemical reactions this will not often be possible. The form (13.9) has thus been used extensively, although (13.10) seems logically preferable, especially for reactions in non-ideal solutions. Equations (13.9) differ from those usually given in chemical textbooks only by the inclusion of the factor  $c^{1-i}$ . This is merely a result of our using equilibrium constants relating to atomic fractions rather than concentrations, i.e. of a choice of a different standard state (the pure components).

In many reactions, the entropy of activation does not vary greatly, and hence the rate of reaction appears to be governed by the heat of activation. This differs only slightly from the experimental activation energy. The entropy of activation may be positive or negative; a high positive entropy implies a high probability of formation of an activated complex,

<sup>†</sup> Since there is equilibrium between reactants and activated complexes, the actual free energy difference is zero, of course.

and vice versa. On p. 84 we interpreted  $C_4$  as giving the frequency with which the unit step can be attempted. In many gaseous reactions, it is found that  $C_4$  is approximately equal to the collision frequency, but in some reactions the two differ by many powers of ten. This is because the collision theory fails to take into account the distribution of energy between the different degrees of freedom. A specific example is provided by two polyatomic molecules which combine together to form a single molecule. During the reaction, three translational and three rotational degrees of freedom disappear and are replaced by six new vibrational degrees of freedom. There is thus a relatively small probability of transition, even when the total energy is available, unless the coordinated motions happen to be suitably related. This is merely another way of saying that  $\Delta S^*$  is negative. Conversely, for a unit step in which a relatively highly organized form of energy becomes more randomly distributed,  $\Delta S^*$  is positive and there is a high probability of transition.

We now have to correlate the experimental activation energy with the results of reaction rate theory.  $E$  is determined experimentally from the equation

$$\varepsilon/k = E/R = -d(\ln k)/d(1/T) = T^2 d(\ln k)/dT.$$

In some textbooks, (13.8) is used, so that

$$\frac{E}{R} = T + T^2 \frac{\partial}{\partial T} \left\{ \ln \frac{K^*}{c^{i-1}} \right\} + T^2 \frac{\partial}{\partial T} \left\{ \ln \frac{\gamma_A \gamma_B \dots}{\gamma^*} \right\}.$$

Now

$$\frac{\partial}{\partial T} \left\{ \ln \frac{K^*}{c^{i-1}} \right\} = \frac{\partial}{\partial T} (\ln K_c^*) = \frac{\Delta U^*}{RT^2},$$

where  $\Delta U^*$  is the internal energy of activation per mole for the unit step reaction. If we neglect the variation of the activity coefficients with temperature, we have

$$E = RT + \Delta U^* = RT + \Delta H^* - p \Delta v^*, \quad (13.11)$$

where  $\Delta v^*$  is the change in volume in forming the activated complex and may be neglected for condensed assemblies. Since the  $RT$  term is negligible in comparison with  $E$  the experimental activation energy is very nearly equal to the theoretical "activation enthalpy" in condensed assemblies.

This treatment is obviously unsatisfactory, and the Zener method (13.10) leads to better results. In cases where it may be applied, e.g. in solids, it is usually permissible to write  $Q_v$  as  $kT/h\nu$ , where  $\nu$  is a characteristic vibration frequency for the reaction concerned. Equation (13.10) thus becomes

$$k_f = \nu \exp(-\Delta_a G/RT) = \nu \exp(\Delta_a S/R) \exp(-\Delta_a H/RT) \quad (13.12)$$

and differentiation gives directly

$$E = \Delta_a H, \quad (13.13)$$

so that the physical quantity in the rate equation given by Zener can not only be defined more clearly, but also has a clear identification with the experimental measurement.

We now discuss the difficulty mentioned on p. 80, namely that in the solid state, the reactants necessarily interact with all the other atoms present in forming the activated complexes. The reactant is frequently a single atom, as in most diffusion mechanisms for example, and the approximation involved is thus that of treating a many body problem as a single body problem. Vineyard (1957) developed a more general treatment which is based only on the rate theory assumptions that well-defined initial and saddle-point configurations exist in quasi-equilibrium, and on the additional assumption that the motions of atoms in both configurations may be treated as small, simple, harmonic oscillations. We abandon the rather artificial chemical description in terms of concentrations of reacting species and consider instead a crystal of  $N$  atoms with  $3N$  degrees of freedom represented by spatial coordinates  $x_1, \dots, x_{3N}$ . Let the potential energy of the crystal be  $U(y_1, \dots, y_{3N})$ , where  $y_1 = m_1^{1/2} x_1$ , etc., and  $m_i$  are the masses associated with each  $x_i$ . Then in a  $3N$ -dimensional configurational space  $y_i$ ,  $U$  has a minimum at some initial and final configurations ( $A$  and  $P$ ) and there is also a critical configuration  $M^*$  for the transition  $A \rightarrow P$ . A hypersurface  $S$  of dimensionality  $3N-1$  may be defined so that it passes through  $M^*$  and is perpendicular to the contours of  $U$  elsewhere. A transition from  $A$  to  $P$  occurs if a system crosses  $S$  with finite velocity.

By arguments essentially similar to those developed above for the Eyring theory, Vineyard showed that the rate constant for the transition  $A \rightarrow P$  is given by

$$k_f = \left(\frac{kT}{2\pi}\right)^{1/2} \frac{\int_S \exp(-U/kT) dS}{\int_V \exp(-U/kT) dV}, \quad (13.14)$$

where the upper integration is over the hypersurface  $S$  and the lower integral is over the region of configuration space bounded by  $S$ . This equation may be compared with (13.1); it contains the ratio of two configurational partition function integrals, which must be calculated so as to include all atoms and degrees of freedom.<sup>†</sup> The theory of small vibrations is now used to expand  $U$  as a Taylor series about its values  $U(A)$  and  $U(M^*)$  respectively. By means of an axis transformation, the normal coordinates  $q_1, \dots, q_{3N}$  and frequencies  $\nu_1, \dots, \nu_{3N}$  for the initial state  $A$  are introduced, and the value of  $U$  in the volume integral becomes

$$U \simeq U(A) + \frac{1}{2} \sum_{i=1}^{3N} (2\pi \nu_i q_i)^2. \quad (13.15)$$

Hence the partition function for the initial state may be written

$$\int_V \exp(-U/kT) dV = \prod_{i=1}^{3N} (kT/2\pi \nu_i^2) \exp\{-U(A)/kT\}. \quad (13.16)$$

Similarly, the normal coordinates  $q'_i$  and frequencies  $\nu'_i$  for vibrations about  $M^*$  but confined to the hypersurface  $S$  give the approximation for the activated complex

$$U \simeq U(M^*) + \frac{1}{2} \sum_{i=1}^{3N-1} (2\pi \nu'_i q'_i)^2 \quad (13.16a)$$

<sup>†</sup> Note that the partition function integrals in (13.14) have dimensions of area and volume in the hyper-space  $y_i$ ; i.e. the ratio of the two integrals has dimension  $(\text{mass})^{-1/2} (\text{length})^{-1}$ .

and

$$\int_S \exp(-U/kT) dS = \prod_{i=1}^{3N-1} (kT/2\pi \nu_i')^2 \exp(-U(M^*)/kT). \quad (13.17)$$

Thus, finally, from (13.14)

$$k_f = \nu^* \exp - \{ [U(M^*) - U(A)] / kT \}, \quad (13.18)$$

where the effective frequency is

$$\nu^* = \left( \prod_{i=1}^{3N} \nu_i \right) / \left( \prod_{i=1}^{3N-1} \nu_i' \right). \quad (13.19)$$

The frequency  $\nu^*$  is thus the ratio of the product of the  $3N$  normal frequencies of the assembly in the initial configuration to the  $3N-1$  normal frequencies of the assembly constrained in the saddle point configuration. In general the frequencies  $\nu_i'$  will differ from the corresponding frequencies  $\nu_i$ , and it will not be justifiable to identify  $\nu^*$  with, for example, the best Einstein or Debye frequency obtained from specific heat data.

Vineyard's expression may also be written in the form of eqn. (13.12) by a device similar to that used by Wert and Zener. Consider a hypersurface  $S_0$  which is similar to  $S$  and passes through the point  $A$  of configuration space so that its normal at  $A$  is along the line of force leading to  $M^*$ . By defining a new partition function for vibrations restricted to the hypersurface  $S_0$ , Vineyard showed that (13.12) is valid with

$$\nu^* = \nu \exp(\Delta S/k), \quad (13.20)$$

$$\Delta S = k \ln \left\{ \left( \prod_{i=1}^{3N-1} \nu_i^0 \right) / \left( \prod_{i=1}^{3N-1} \nu_i' \right) \right\}, \quad (13.21)$$

where  $\nu_i^0$  are the normal frequencies for the assembly constrained to lie in  $S_0$ . The frequency  $\nu$  is thus given by

$$\nu = \left( \prod_{i=1}^{3N} \nu_i \right) / \left( \prod_{i=1}^{3N-1} \nu_i^0 \right) \quad (13.22)$$

and corresponds approximately to its previous description as the frequency of vibration in the initial state along the reaction path (normal to  $S_0$ ). The interpretation of  $\Delta_a G$ ,  $\Delta_a H$ ,  $\Delta_a S$  is also similar to that given above, the assembly being constrained not to vibrate except in a hypersurface which is everywhere normal to the path from  $A$  to  $P$  in the  $3N$  dimensional configuration space.

When reaction rate theory is applied to diffusion it is important to consider the effects of using different isotopes as tracer elements. The activation energies may then be assumed to be almost equal, but the frequency factors differ because of the effects of the isotopic mass on the normal modes. The simplest assumption is that the migrating mechanism involves only one atom, and that there is little or no coupling between this atom and the rest of the crystal. If the rates of migration are then calculated for two isotopes, all frequencies cancel with the exception of those localized frequencies which are associated with the migrating atoms. Thus for isotopes  $A$  and  $\alpha$ ,  $\nu_{iA}/\nu_{i\alpha} = \nu'_{iA}/\nu'_{i\alpha}$  and

$$k_A/k_\alpha = \nu_A/\nu_\alpha = (m_\alpha/m_A)^{1/2}, \quad (13.23)$$

where  $\nu_A$  and  $\nu_\alpha$  are corresponding frequencies of vibration for the two isotopes, and the result follows since atomic frequencies are inversely proportional to the square roots of the masses.

The assumption that the migrating atom is effectively decoupled from the rest of the crystal is rather extreme. Normally there will be a change of volume in the activated state, and the displacements of the surrounding atoms then produce changes in the vibrational modes. Vineyard treated this complication by defining an effective mass for the migrating atom; a further development due to Le Claire (1966) shows that the ratio of rate constants can be expressed in the form

$$k_A/k_\alpha = \nu'_{1A}/\nu'_{1\alpha}, \quad (13.24)$$

where  $\nu'_{1A}$  is the frequency of the unstable mode normal to the hypersurface  $S_0$ . Introducing

$$\Delta K = \{(\nu'_{1A}/\nu'_{1\alpha}) - 1\} / \{(m_\alpha/m_A)^{1/2} - 1\}, \quad (13.25)$$

Le Claire showed that  $\Delta K$  is that part of the kinetic energy in the mode of frequency  $\nu'_1$  which is possessed by the migrating atom, divided by the total kinetic energy in this mode. Equation (13.27) was first derived by Mullen (1961), who defined  $\Delta K$  as the fraction of the translational kinetic energy which is possessed by the migrating atom as it passes over the saddle point; Le Claire's derivation shows that only the energy in the mode  $\nu'_1$  is to be taken into account in defining this fraction.

Quantum effects are neglected in all the above derivations but may become important in certain circumstances, e.g. in thermally activated processes below the Debye temperature  $\theta_D$ . If the classical partition functions are replaced by quantum partition functions, eqn. (13.19) is modified to

$$v^* = \frac{\prod_{i=2}^{3N} [\exp(-\frac{1}{2}h\nu'_i/kT) / \{1 - \exp(-\frac{1}{2}h\nu'_i/kT)\}]}{\prod_{i=1}^{3N} [\exp(-\frac{1}{2}h\nu_i/kT) / \{1 - \exp(-\frac{1}{2}h\nu_i/kT)\}]} \frac{kT}{h}. \quad (13.26)$$

Taking the first term of the expansion

$$e^{-z}/(1 - e^{-z}) = \frac{1}{2} \operatorname{cosech} z = \left(\frac{1}{2z}\right) \left(1 - \frac{z^2}{6} + \dots\right)$$

we obtain (13.19), whilst if the first two terms are retained,

$$v^* = \frac{\prod_{i=1}^{3N} \nu_i}{\prod_{i=2}^{3N} \nu'_i} \left\{ 1 - \frac{1}{24} \left(\frac{h}{kT}\right)^2 \left(\sum_2^{3N} (\nu_i)^2 - \sum_1^{3N} (\nu_i)^2\right) \right\}. \quad (13.27)$$

Le Claire shows that this equation leads to a revised form of (13.23):

$$\begin{aligned} \frac{k_A}{k_\alpha} - 1 &= \left[ \left(\frac{m_\alpha}{m_A}\right)^{1/2} - 1 \right] \left\{ 1 - \frac{1}{12} \left[ \frac{m_\alpha}{m_A} + \left(\frac{m_\alpha}{m_A}\right)^{1/2} \right] \left(\frac{h\nu_\alpha}{kT}\right)^2 \left[ \left(\frac{\nu'_\alpha}{\nu_\alpha}\right)^2 - \frac{3}{2} \right] \right\} \\ &\simeq \left[ \left(\frac{m_\alpha}{m_A}\right)^{1/2} - 1 \right] \left\{ 1 - \frac{1}{6} \left(\frac{h\nu}{kT}\right)^2 \left[ \left(\frac{\nu'}{\nu}\right)^2 - \frac{3}{2} \right] \right\} \end{aligned} \quad (13.28)$$

when  $m_A \approx m_\alpha$  so that  $v_A \approx v_\alpha \approx v$ . For numerical estimates,  $h\nu/k$  is replaced by the Debye temperature  $\theta_D$ , and Le Claire concludes that for typical diffusion measurements the maximum correction, represented by the term in braces in (13.28), is likely to be only a few per cent except perhaps in alkali metals.

Vineyard's theory gives the best description currently available for activated processes in the solid state. We may also, however, briefly mention the dynamical approach to rate theory, which has been developed for the particular case of diffusion by Rice and his collaborators (Rice, 1958; Rice and Nachtrieb, 1959). In an attempt to avoid the difficulty mentioned on p. 81, this theory considers explicitly the probability that a diffusing atom will have sufficient amplitude of vibration to move in a given direction, and that the neighbouring atoms will move out of the way. The parameters which define this model are shown by Rice to be related to the normal coordinates of the crystal, a normal mode vibrational analysis which includes the effect of lattice defects being supposed to have been made. The details of this theory are rather complex, and will not be discussed here since the validity of the approach is not generally accepted. The model does not seem to be amenable to an exact calculation, except under rather restrictive conditions, for example that the atoms are harmonically bound. With such assumptions, the expression for the rate constant may be put into a form equivalent to that given by the chemical rate theory, although the physical interpretations of the terms are slightly different.

We have already emphasized that a dynamical theory may be superior in principle to a thermodynamic theory, since the assumption of an activated state in equilibrium with the reactants is a marked over-simplification. The activated state has so short a lifetime (at most a few vibrations) that the assumptions of the chemical theory are at best unverifiable. On the other hand, the dynamical theory as developed at present is forced to make other assumptions, which are also rather unrealistic.

#### REFERENCES

- CRUSSARD, C. (1948) *Compt. Rend. Semaine d'Études de la Physique des Métaux*, 39; (1952) *L'État solide* (Report 9th Solvay Conference Brussels), 346.
- GLASSTONE, S., LAIDLER, K. J., and EYRING, H. (1941) *The Theory of Rate Processes*, McGraw-Hill, New York.
- LE CLAIRE, A. D. (1966) *Phil. Mag.* **14**, 1271.
- MULLEN, J. G. (1961) *Phys. Rev.* **121**, 1649
- RICE, S. A. (1958) *Phys. Rev.* **112**, 804.
- RICE, S. A. and NACHTRIEB, N. H. (1959) *J. Chem. Phys.* **31**, 139
- VINEYARD, G. H. (1957) *J. Phys. Chem. Solids* **3**, 121.
- WERT, C. A. and ZENER, C. (1949) *Phys. Rev.* **76**, 1169.
- ZENER, C. (1952) *Imperfections in Nearly Perfect Crystals*, p. 305, Wiley, New York.

## CHAPTER 4

# *The Thermodynamics of Irreversible Processes*

### 14. MICROSCOPICAL REVERSIBILITY: THE ONSAGER RECIPROCAL RELATIONS

Classical thermodynamics is concerned primarily with the interdependence of certain well-defined macroscopic concepts (temperature, pressure, entropy, energy, composition, etc.) possessed by a closed assembly. The usual thermodynamic equations are valid only for assemblies at equilibrium and for reversible transitions between such equilibrated assemblies. When thermodynamic considerations are applied to irreversible (i.e. "natural") processes, the equations become inequalities, and are much less useful. For example, the principle of the increase in entropy during an adiabatic irreversible process provides information only about the direction of the change.

The theory of transformations is a description of a particular class of irreversible processes. The details of such a description, as emphasized in Chapter 1, depend on assumptions about atomic interactions and the nature of the elementary steps of the reaction. This type of theory will be our main concern, and may appropriately be described as kinetic. Kinetic theory is not, however, confined to the non-equilibrium state; thermodynamic equilibrium is characterized by stationary values of the macroscopic parameters, but each of these is effectively the average value of some continuously varying microscopic concept. The connection between the microscopic properties of the systems of the assembly and the macroscopic (measurable) properties is made by statistical mechanics. The second law of thermodynamics then follows from a single basic postulate, and may be regarded as the necessary consequence of the averaging process applied to an assembly which has reached dynamic equilibrium.

Following this approach, it is natural to enquire whether there is any statistical principle which when applied to the kinetic laws of an irreversible process also imposes restrictions on the macroscopic parameters of the assembly. In fact, such restrictions have been proposed for certain kinds of change which can be described as steady state phenomena, and the attempt to produce a formal theory of this development has become known as the thermodynamics of irreversible processes. Although this theory has been discussed mainly in the last 30 years, the one useful and reasonably well-accepted result was obtained by Onsager in 1931. It is as well to emphasize here that much work in this field is highly controversial, and almost all the "results" or "principles" enunciated by some workers are thought in-

valid by others. We shall be mainly concerned with the Onsager relations, which are widely used, but even here we shall see that the usual way of applying the result rests on doubtful foundations.

A steady state implies that at each part of the assembly, which is now "open" rather than "closed", the thermodynamic quantities are time independent, even though energy dissipation is occurring. An example of such a process is the conduction of heat down a bar, one end of which is maintained at a uniform high temperature. The theory has, in fact, been applied mainly to transport phenomena in which there is a flow of heat, matter, electricity, etc. In the study of transformations, the theory is most important in the treatment of diffusion in the solid state, but attempts have also been made to use the theory directly in more complex phenomena, such as the migration of an interface. The thermodynamics of irreversible processes has also been applied to chemical reactions, though only in the limiting case of very close approach to equilibrium.

The macroscopic laws describing a steady-state process are often known or assumed to require a linear relation between the rate at which the process occurs and the "thermodynamic force" which causes the process. The most familiar example is Ohm's law, which states there is a linear relation between the flow of electrical current and the potential gradient or electromotive force. In general, the flow rates may be called currents (diffusion currents, heat currents, etc.) and are linearly related to the forces or affinities. The forces are the gradients of some potential function (temperature, chemical potential, etc.) which may be regarded as responsible for the currents.

The terms forces and currents used in this thermodynamic sense are rather misleading. For example, although both these quantities will usually be vectors, as in transport phenomena, they may in principle be tensors of any order. An example of scalar forces and currents occurs in chemical reactions; the forces are proportional to the differences in chemical potential, and the currents to the corresponding rates of reaction.

The assumption that a current  $I$  and its conjugate force  $Z$  are linearly related may be expressed

$$I = MZ \text{ (vector quantities)} \quad \text{or} \quad I = MZ \text{ (scalar quantities)}. \quad (14.1)$$

In this equation we have assumed that when the force and current are both vectors they are parallel to each other. The scalar quantity  $M$  is given different names in different steady-state phenomena, but is always of the nature of a conductance, since its reciprocal expresses the resistance to flow.

Equation (14.1) will be satisfied in isotropic media, e.g. in randomly orientated polycrystalline solids, but in single crystals it is necessary to introduce a more general relation

$$I_i = M_{ij}Z_j \quad (i, j = 1, 2, 3) \quad (14.2)$$

or, in matrix form,

$$i = M Z,$$

where  $M_{ij}$  form the components of a second-order tensor, so that we now have linear relations between the components of the currents and the components of the conjugate forces.

The situation of interest in the thermodynamic theory arises either when we have a three-dimensional transport process (as in eqn. (14.2)) or when we have two or more currents and forces of different kinds occurring together. There may then be mutual coupling or interference between the separate effects. For example, if two metallic junctions are maintained at different temperatures, there will be a flow of electricity as well as of heat, and conversely if there is an electrical potential difference maintained between them, heat is absorbed or liberated at the junctions.

In general, we may suppose that we have different kinds of vector flow  $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n$  and different kinds of vector force  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ . If the currents and their conjugate forces are parallel, as in eqn. (14.1), we may write for the interaction of all these quantities

$$\mathbf{I}_i = M_{ij}\mathbf{Z}_j \quad (i, j = 1, 2, \dots, n), \quad (14.3a)$$

where the diagonal elements of the matrix  $M$  are the conjugate conductances, and the non-diagonal elements are the interference or coupling conductances. Similarly if the flows are scalars we have

$$I_i = M_{ij}Z_j \quad (i, j = 1, 2, \dots, n). \quad (14.3b)$$

Note that we assume all forces to be either scalars or vectors; there is, of course, non interaction between currents and forces of different tensor order.

If there is a tensor relation between each current and the conjugate force, we expect the *component* of each current to be a linear function of all the components of each force. We may write this equation in the form

$$I_i = M_{ij}Z_j \quad (i, j = 1, 2, \dots, 3n) \quad (14.3c)$$

if we interpret each subscript,  $i$  or  $j$ , as giving a double specification—the type of flow and the axis to which the vector component refers. This notation is similar to that used for the Hooke's law eqns. (11.15), and is useful because it preserves the relation between currents and forces in square matrix form. The diagonal elements of the matrix are now the conductances relating the component of any flow along one of the axes with the component of its conjugate force along the same axis; the non-diagonal elements relate the component of any flow along one axis with the components of its conjugate force along the other axes, or with the components of the non-conjugate forces along any axis. Clearly, we have reduced the vector problem to a scalar representation, in which component flows and forces are conjugate to each other.

Equations (14.3a, b, or c) are called the phenomenological equations or the thermodynamic equations of motion. We emphasize here that thermodynamics can say nothing of their validity; this must be determined either by experiment or by arguments based on kinetic theory.

The important result on which the thermodynamics of irreversible processes largely depends was first obtained by Onsager (1931a, b). This states that provided a proper choice has been made of conjugate currents and forces, the matrix  $M$  must be symmetrical and

$$M_{ij} = M_{ji}. \quad (14.4)$$

We thus see that the theory only yields useful results when there is mutual interference of effects. The proper choice of currents and forces is considered below; this aspect of the theory can lead to real difficulties.

We shall not attempt to give a full derivation of Onsager's relations, as this would mean a lengthy diversion into statistical theory. The basis of the equations lies in the principle of microscopical reversibility, which states that under equilibrium conditions, any microscopic (i.e. molecular) process and its reverse take place on the average at the same rate. Expressed mathematically, this means that the mechanical equations of motion of individual particles are symmetric with respect to the time, so that the transformation  $t \rightarrow -t$  leaves them unchanged. This result is founded on quantum statistics and is of very general validity; the only exceptions are effects in non-conservative force fields (electromagnetic induction, Coriolis forces), in which the direction of the field has to be reversed to obtain an invariant transformation from  $t$  to  $-t$ . We have already noted that the chemical principle of detailed balancing is a special case of the above principle.

Onsager obtained the reciprocal relations by considering fluctuations in an assembly at equilibrium, and relating these to macroscopic flows. Suppose we measure these fluctuations by a set of variables<sup>†</sup>  $e_i$  which represent the local deviations in pressure, temperature, density, etc., from the mean or equilibrium values of these quantities. Thus the mean value of  $e_i$  over a large region, or for a small region over an appreciable period of time, is zero. Now if at time  $t$  the local deviation of type  $i$  in any region is  $e_i$ , we may use the symbol  $\overline{e_i(t)e_j(t+\tau)}$  to represent the average probability of a fluctuation  $e_i$  being followed by a fluctuation  $e_j$  at a time  $\tau$  later. According to the principle of microscopical reversibility, this sequence of two fluctuations must occur just as frequently as its reverse, so that

$$\overline{e_i(t)e_j(t+\tau)} = \overline{e_j(t)e_i(t+\tau)}.$$

We now consider how the averages are to be evaluated. If we first take a fixed fluctuation at time  $t$ , then at time  $t+\tau$  there will in general be an unpredictable series of deviations. We may, however, take the average value of  $e_j(t+\tau)$  by considering a large number of situations in which the fluctuation at  $t$  is fixed, and finding the mean of the values of  $e_j$  after a time interval  $\tau$ . We denote this average by  $\overline{e_j(t+\tau)}_t$ , the square brackets indicating that the situation at time  $t$  is fixed whilst the average is evaluated. We can then obtain the average value of  $e_i(t)e_j(t+\tau)$  by allowing the fixed fluctuation at  $t$  to vary and averaging over these variations. We thus have

$$\overline{e_i(t) \overline{e_j(t+\tau)}_t} = \overline{e_j(t) \overline{e_i(t+\tau)}_t}.$$

Now subtracting from each side the average product of  $e_i(t)$  and  $e_j(t)$ , which may be written  $\overline{e_i(t)e_j(t)} = \overline{e_i(t) \overline{e_j(t)}_t} = \overline{e_j(t) \overline{e_i(t)}_t}$ ,

$$\overline{e_i(t) \overline{e_j(t+\tau) - e_j(t)}_t} = \overline{e_j(t) \overline{e_i(t+\tau) - e_i(t)}_t}. \quad (14.5)$$

<sup>†</sup> The symbol  $e$  has this meaning only in this chapter.

Onsager now makes an important assumption, namely that the average rates of change of the quantities  $e_i$  with respect to time are equal to the corresponding macroscopic flows. The fluctuations are thus supposed to be small transient flows of heat, matter, etc. This type of assumption is implicit in the treatment of such fluctuation phenomena as Brownian motion, and may be shown to be highly probable by kinetic arguments. We may thus write

$$de_i/dt = (1/\tau) \overline{[e_i(t+\tau) - e_i(t)]} = I_i = M_{ik}Z_k$$

and  $de_j/dt = I_j = M_{jk}Z_k$ . (14.6)

Substituting into (14.5),

$$\overline{e_i(t) [M_{jk}Z_k]_t} = \overline{e_j(t) [M_{ik}Z_k]_t}. \quad (14.7)$$

Now let us consider the deviation in the entropy associated with a fluctuation specified by  $e_1, e_2, \dots$ . Since  $S$  is a maximum in the equilibrium state ( $\Delta S = 0$ ), we may express the entropy fluctuation  $\Delta S$  as a quadratic function of the state parameters  $e_j$ , so that

$$\Delta S = -\frac{1}{2}g_{ij}e_i e_j, \quad (14.8)$$

where  $g_{ij}$  is a positive definite form. Using Boltzmann statistics, it follows that the probability of a particular fluctuation in which the state parameters (regarded as continuous) have values lying between  $e_1$  and  $e_1 + de_1$ ,  $e_2$  and  $e_2 + de_2$ , etc., is given by

$$P de_1, de_2, \dots, de_n = \frac{\exp(\Delta S/k) de_1 de_2 \dots de_n}{\int_{e_1} \int_{e_2} \dots \int_{e_n} \exp(\Delta S/k) de_1 de_2 \dots de_n}$$

the denominator on the right-hand side merely being a normalizing factor to make the total integral of  $P$  equal to unity. With this equation, it is easily proved that the mean value of the product  $e_i(\partial\Delta S/\partial e_j)$  is given by

$$\overline{e_i(\partial\Delta S/\partial e_j)} = -k\delta_{ij}.$$

Let us make the identification

$$Z_j = (\partial\Delta S/\partial e_j) = -g_{ij}e_i \quad (14.9)$$

so that

$$\overline{e_i Z_j} = -k\delta_{ij}. \quad (14.10)$$

Now substituting into (14.7),

$$-k\delta_{ik}M_{jk} = -k\delta_{jk}M_{ik} \quad \text{or} \quad M_{ji} = M_{ij},$$

which is Onsager's relation. The above identification of  $Z$  will be considered further in the next section; in effect, this is the meaning of a proper choice of conjugate forces.

We note that eqn. (14.4), although founded on a microscopical principle, expresses relations between macroscopic quantities. This situation is analogous to that encountered

in classical thermodynamics since the second law is also a macroscopic law based on a microscopic principle (that of equal *a priori* probability of all microstates). We also note the formal similarity between the result obtained by Onsager for steady-state processes, and the symmetry of the scheme of elastic compliances or stiffnesses, eqn. (11.17). The symmetry of the latter, as does that of other equilibrium properties, depends on the assumption of an energy function; in a sense, the principle of microscopical reversibility is a variational form of this assumption for non-static processes.

## 15. ENTROPY PRODUCTION IN NATURAL PROCESSES

Thermodynamic quantities are strictly defined only for assemblies which are in equilibrium. We begin this section with a discussion of the circumstances under which it is meaningful to assign thermodynamic parameters to non-equilibrium assemblies. Consider an irreversible transformation in an assembly which is initially and finally at equilibrium. The change in entropy is then defined precisely in terms of other functions which describe the equilibrium states, and the second law tells us that entropy is created during the process. Now if it is possible to know the local pressure, volume, temperature, chemical composition, etc., at each stage in the transformation to a sufficient degree of accuracy, the entropy is also defined, at least to within narrow limits. The local values of the macroscopic parameters are obtained in principle by isolating small regions of the assembly (which still contain large numbers of molecules) and allowing each to come to equilibrium. The procedure is valid only when the deviations from equilibrium are everywhere small, and it is only under these conditions that we are justified in thinking of the assembly as possessing macroscopic properties.

In most processes, the difficulty of defining local parameters arises mainly from the lack of equilibrium between the different forms of energy which are present. The temperature is determined by the mean value of the kinetic energy, and has meaning only when there are sufficient collisions or other molecular interactions for the excited forms of energy to attain local equilibrium very rapidly. This means that the free path for energy transfer must be much smaller than the distances over which the temperature varies, and is another way of stating that the situation must not be too far removed from equilibrium.

The second law shows that entropy, unlike energy, is not conserved during natural changes. During a reversible change in which an amount of heat  $dq$  enters the assembly from the surroundings, the entropy of the assembly increases by  $dS$ , where

$$T dS = dq = -T d_e S$$

and  $d_e S$  is the change in entropy of the environment. More generally, for any reversible change involving flows of energy, matter, etc., we have

$$dS + d_e S = 0,$$

and for an irreversible change

$$dS + d_e S > 0.$$

We may define  $d_i S$  as the irreversible entropy which is *created* during the process, where

$$d_i S = dS + d_e S. \quad (15.1)$$

Note that  $dS$  and  $d_e S$  may have either sign, but  $d_i S$  must be positive or zero.

There is a particularly simple expression for  $d_i S$  if we consider irreversible processes at constant temperature and pressure. Then we have

$$dG = dU + p dV - T dS = -T d_e S - T dS,$$

so that

$$d_i S = -dG/T,$$

where  $G$  is the Gibbs function for the assembly. The rate at which entropy is produced during the process is then

$$d_i S/dt = -(1/T) dG/dt. \quad (15.2)$$

This quantity may be called the entropy source strength.

More generally, the above discussion shows that we may write the increment of entropy during any process which is never too far removed from equilibrium as

$$T dS = dU + p dV - g_i dN_i, \quad (15.3)$$

where  $g_i$ ,  $N_i$  represent the chemical potential per molecule and the number of molecules of the  $i$ th substance within the assembly at any time during the change. If we use this equation for  $dS$  and subtract from it the contribution to the change in entropy of the assembly produced by the transfer of work, heat, matter, etc., from the surroundings, we obtain the irreversible entropy produced,  $d_i S$ .

We may now use Onsager's identification of the macroscopic flows with fluctuation phenomena, eqn. (14.6), to relate the rate of entropy production to the forces and fluxes of the last section. The quantity  $d_i S/dt$  of this section is to be identified with  $d\Delta S/dt$  for a fluctuation. Using eqns. (14.8) and (14.9), we find that we can express the entropy source strength in the form

$$d_i S/dt = I_k Z_k, \quad (15.4)$$

where  $I_k$ ,  $Z_k$  are conjugate currents and thermodynamic forces, and, as explained above, may be either scalars or parallel vector components.

As an example, we consider a single chemical reaction at constant temperature and pressure. From either (15.2) or (15.3)

$$d_i S/dt = -(1/T) g_i (dN_i/dt).$$

Suppose the reaction is written

$$a_i A_i = 0,$$

where  $a_i$  represents any of the quantities  $a$ ,  $b$ ,  $p$ ,  $q$ ,  $\dots$ , and  $A_i$  any of the components  $A$ ,  $B$ ,  $\dots$ ,  $P$ ,  $Q$ ,  $\dots$ , in eqn. (12.1). If there were initially  $N_{i,0}$  molecules of species  $i$  in

the assembly, we may use the quantity  $\xi = (N_i - N_{i,0})/a_i$  as a measure of the amount of the reaction, and this quantity is moreover the same for all species at a given stage in the reaction. If we adopt the convention of calling  $a_i$  positive for the products of the reaction, then  $\xi$ , called the degree of advancement, is also positive. Substituting  $d\xi = dN_i/a_i$

$$d_i S/dt = -(1/T)(d\xi/dt)a_i g_i. \quad (15.5)$$

The quantity  $-a_i g_i$  may now be called the affinity of the reaction; it is a measure of the free energy difference between reactants and products, and is thus positive. The quantity  $d\xi/dt$  is a measure of the velocity of the reaction. The right-hand side may thus be written as the product of a single thermodynamic force and its associated current, and provides an example of eqn. (15.4) when there is no coupling.

Now suppose we have a set of chemical components amongst which  $r$  different reactions are possible. We write the reactions

$$\begin{aligned} a_{1i} A_i &= 0, \\ a_{2i} A_i &= 0, \\ \dots, \\ a_{ni} A_i &= 0. \end{aligned}$$

The degree of advancement of each reaction may be represented by  $\xi_j = (N_i - N_{i,0})/a_{ji}$  and the rate of this reaction is thus  $d\xi_j/dt = (1/a_{ji}) dN_i/dt$ . This gives

$$d_i S/dt = -(1/T)(d\xi_j/dt)a_{ji} g_i. \quad (15.6)$$

Each quantity  $-a_{ji} g_i$  may now be called the affinity for the reaction  $j$ , and the entropy source strength is thus equal to the sum of the products of the affinities and the corresponding flows, as in eqn. (15.4). Note that the sum of the products of the affinities and reaction rates must be positive, but there is no need for any individual product to be positive. When the reactions are coupled, some of them may be driven "backwards".

We have quoted the example of chemical reactions to show how eqn. (15.4) is often used in phenomenological theories to identify the forces and fluxes of eqns. (14.3). We have now to emphasize that the expression for the entropy source strength, although it appears to be very convenient in phenomenological theories, is in fact a most unsatisfactory way of selecting forces and fluxes. It is implied in many books, e.g. de Groot (1951), that the Onsager relations eqn. (14.4) are valid for all forces and fluxes which simultaneously satisfy eqns. (14.3) and (15.4). Coleman and Truesdell (1960) have shown clearly that this cannot be true, the difficulty being that the fact that (15.4) follows from (14.6) to (14.9) does not ensure the validity of (14.6) to (14.9) given (15.4).

Suppose we have a set of forces and fluxes which satisfy eqns. (14.3), (15.4), and (14.4). Now let  $W$  be any non-zero antisymmetrical matrix with the same number of rows and columns as  $M$ , so that  $W_{ij} = -W_{ji}$ . We can now define new fluxes and forces by the linear transformation

$$\begin{aligned} I_i &= I_i + W_{ij} Z_j, \\ Z_i &= Z_i. \end{aligned} \quad (15.7)$$

Then it follows that

$$I'_i Z'_i = I_i Z_i + W_{ij} Z_i Z_j,$$

and since  $W_{ij} = -W_{ji}$ , the last term is zero and

$$I'_i Z'_i = I_i Z_i = d_i S/dt. \tag{15.8}$$

We also have from (15.7)

$$I'_i = (M_{ij} + W_{ij}) Z_j = M'_{ij} Z'_j,$$

which is equivalent to (14.3) but has new mobilities

$$M'_{ij} = M_{ij} + W_{ij} \tag{15.9}$$

which do not form a symmetric matrix. Thus by beginning with forces and fluxes which satisfy (14.3) and (15.4) and the Onsager relations (14.4), we can construct by a simple linear transformation a new set of forces and fluxes which satisfy (14.3) and (15.4), but which do not satisfy the Onsager relations. Moreover, this argument is reversible in that if we have forces and fluxes for which the Onsager relation is not valid, we can construct new forces and fluxes for which it is valid by splitting the matrix of coefficients  $M'_{ij}$  into symmetrical and antisymmetrical parts, and redefining the fluxes by a transformation which is the inverse of (15.7). It might be thought that this is a satisfactory way of obtaining the correct forces and fluxes which are both physically significant and satisfy the Onsager relations: unfortunately this choice is not unique, since any transformation of form

$$\begin{aligned} Z_i^* &= A_{ik} Z_k \\ I_i^* &= A_{ki}^{-1} I_k, \end{aligned} \tag{15.10}$$

will yield another set of forces and fluxes which satisfy eqns. (14.3), (15.4), and (14.4) in the same way as the first set.

The general conclusion is thus that the Onsager result cannot automatically be applied to a situation in which the choice of forces and fluxes is based on physical intuition and these forces and fluxes are assumed to be linearly related. This does not mean that the Onsager relations are invalid in such a case; if it can be shown that the chosen forces and fluxes satisfy eqns. (14.3) and (14.9) then the above proof is correct.

Coleman and Truesdell point out that the conventional development of the theory of diffusion in multi-component systems involves the choice of fluxes and forces in a way in which it is not obvious that the Onsager relations are valid. This threw rather a blight over diffusion theories for a considerable time although most authors simply assumed the validity of the Onsager relations, and Howard and Lidiard (1964) pointed out that a detailed theory of diffusion includes the Onsager relations as a consequence of the assumptions on which their theory is based. More recently Kirkaldy (1985) and Kirkaldy and Young (1987) have shown that the relations of the type of eqn. (15.4) give valid Onsager relations provided the currents refer to a conserved quantity such as mass, energy or electricity. Thus this long standing difficulty in the theory of diffusion has at last been removed.

Of course, as emphasised above, the thermodynamic theory does not give any information about the validity of the linear relation (14.3). Thus even if suitable forces and fluxes have been found, the linear relation between them is a matter for experiment or for kinetic theory. In particular, there is not generally a linear relation between the affinity of a chemical reac-

tion and the time derivative of the degree of advancement, unless the reaction is very close to equilibrium. This may be seen, from the theory of the last chapter. The forward rate of the reaction is proportional to the factor  $\exp(-\Delta_a g/kT)$ , where  $\Delta_a g$  is the free energy of activation per molecule. The back reaction will have the same activated complex, and will thus have a reaction rate proportional to  $\exp\{(-\Delta_a g + g_i a_i)/kT\}$ , since the activation energy is increased by the difference in the standard free energies of reactants and products. The net rate of reaction will thus be proportional to

$$\exp(-\Delta_a g/kT) [1 - \exp(g_i a_i/kT)] \quad (15.11)$$

and will thus be linear in  $-g_i a_i$  only if the reaction is so close to equilibrium that  $-g_i a_i \ll kT$ .

It may be shown quite generally that the validity of the assumption of linear relations and the validity of the proof of Onsager's theorem both depend on the existence of small and definable deviations from equilibrium. In chemical reactions, the free energy per molecule released during the process must be less than  $kT$ . Transport phenomena are usually observed in the linear region; chemical reactions seldom are.

It follows from our discussion above that there is a general class of linear transformations of forces and currents which leaves the reciprocal relations valid. In particular, it is possible to find a linear transformation which will reduce the matrix  $M$  to diagonal form. This means that by suitably combining the forces and currents, we obtain phenomenological relations without cross terms. Such combinations have no particular physical significance.

We note that for any single physical property of a three-dimensional crystal, e.g. properties of the type (14.3c), the Onsager relations must be combined with the symmetry properties of the crystal to obtain the scheme of coefficients. Thus we find that for vector-vector properties, the second order tensor reduces to a scalar if the symmetry is cubic. Properties like the electrical conductivity and the coefficient of diffusion are thus isotropic in cubic materials, although three coefficients are needed to specify properties like the elastic stiffnesses which relate two second order tensors. A detailed survey of the effect of crystal anisotropy on physical properties is given by Nye (1957).

The description which we have given of the thermodynamics of irreversible processes has been concerned almost exclusively with the Onsager relations. In equilibrium thermodynamics, the equilibrium state is characterized by stationary values of certain extensive thermodynamic parameters, the appropriate condition (maximum entropy, minimum free energy, etc.) depending on the external constraints. The discovery of the Onsager relations suggests the possibility of finding similar conditions which will define the steady state, and such a generally applicable principle would undoubtedly be of great assistance in formulating kinetic theories of steady state or quasi-steady-state processes. In fact, a variational principle has been suggested for this purpose by de Groot (1951), following a theorem due to Prigogine, and has been much discussed by these and other workers. This principle is that the rate of entropy production,  $d_i S/dt$ , has a minimum value in the stationary state, consistent with the auxiliary imposed conditions which can be regarded as fixing certain of the forces  $Z_i$  at constant levels.

The principle of minimum entropy production is one of the most controversial developments of the formal theory of this chapter, and we shall not discuss it further here. In recent

years, some workers have used the principle in theories of metallurgical growth processes. However, it is evident that the theory does not hold in all circumstances, as may be shown by choosing particular examples where the answer is known, and no general rule for defining the limits of applicability seems to have been discovered. It thus seems unjustified to apply the theory to the rather complex conditions of a growth process, where it may or may not be valid.

#### REFERENCES

- COLEMAN, B. D. and TRUESDELL, C. (1960) *J. Chem. Phys.* **33**, 28.  
DE GROOT, S R. (1951) *Thermodynamics of Irreversible Processes*, North-Holland, Amsterdam.  
HOWARD, R. E. and LIDIARD, A. B. (1964) *Reports on Progress Phys.* **27**, 161.  
KIRKALDY, J.S. (1985) *Scripta. Met.* **19**, 1307.  
KIRKALDY, J.S. and YOUNG, D.J. (1987) *Diffusion in the Condensed State*, the Institute of Metals, London.  
NYE, J. F. (1957) *Physical Properties of Crystals*, Clarendon Press, Oxford.  
ONSAGER, L. (1931a) *Phys. Rev.* **37**, 405; (1931b) *Ibid.* **38**, 2265.

## CHAPTER 6

### *Solid Solutions*

#### 22. PAIR PROBABILITY FUNCTIONS: THERMODYNAMIC PROPERTIES

A solid phase containing two or more kinds of atom, the relative proportions of which may be varied within limits, is described as a solid solution. Terminal solid solutions are based on the structures of the component metals; intermediate solid solutions may have structures which are different from any of those of the constituents. Most solid solutions are of the substitutional type, in which the different atoms are distributed over one or more sets of common sites, and may interchange positions on the sites. In interstitial solutions, the solute atoms occupy sites in the spaces between the positions of the atoms of the solvent metal; this can only happen when the solute atoms are much smaller than the atoms of the solvent.

We must also distinguish between ordered and disordered solid solutions. In the fully ordered state each set of atoms occupies one set of positions, so that the atomic arrangement is similar to that of a compound. This is only possible at compositions where the ratios of the numbers of atoms of different kinds are small integral numbers, but the atomic arrangement may still be predominantly ordered in this way for alloys of arbitrary composition. In disordered solid solutions, the atoms are distributed among the sites they occupy in a nearly random manner. This classification is only approximate, and we shall formulate these concepts more precisely.

The definition of the unit cell, and the concept of the translational periodicity of the lattice, lose their strict validity when applied to a disordered solid solution. The mean positions of the atoms, considered as mathematical points, will no longer be specified exactly by (5.8), since there will be local distortions depending on the details of the local configurations. Moreover, a knowledge of the type of atom at one end of a given interatomic vector no longer implies knowledge of the atom at the other end, as it does for a pure component or a fully ordered structure. In a solid solution, precise statements of this nature have to be replaced by statements in terms of the probability of the atom being of a certain type.

For many purposes, the strict non-periodicity of the structure is not important, since most physical properties are averages over reasonably large numbers of atoms. Thus the positions of X-ray diffraction maxima depend only on the average unit cell dimensions,

and their intensities depend only on the mean concentrations of atoms of different kinds and the mean interatomic distances. An approximate description in which the structure is regarded as having a unit cell of fixed size, with atomic positions occupied by identical scattering centres of averaged atomic scattering factor, thus suffices for a description of the main features of the X-ray diffraction pattern. A lattice vector of this structure may actually connect two unlike atoms, but is regarded as connecting two average atoms.

If two metals have the same crystal structure, they may form a single solid solution, and the lattice parameter then varies continuously with composition from the value characteristic of one pure metal to that of the other. In some assemblies, the edge length of the unit cell, or the volume per atom, is approximately linear with the atomic fraction of solute; this is described as Vegard's law. The work of Hume-Rothery and his collaborators (see Hume-Rothery and Raynor, 1954) has shown that a necessary (but not sufficient) condition for the formation of extensive terminal solutions is that the sizes of the atoms in the pure components shall not differ by more than  $\sim 15\%$ . This is called the size factor rule.

The localized distortion around a solute atom cannot be observed easily, but it may be studied by X-ray diffraction techniques in which the diffuse background scattering is measured. This type of X-ray measurement also enables us to deduce the probability of a given interatomic vector connecting atoms of given types. In theoretical treatments of the properties of solid solutions, it is usual to treat all atoms as situated on the sites obtained from the mean lattice, and this approximation is adequate in view of the severe limitations of the theories in other respects. The way in which the sites are occupied may then be described by a set of probability factors.

Consider a solid solution containing atomic fractions  $x_A, x_B, x_C, \dots$ , of components  $A, B, C, \dots$ , all of which may occupy any of the atomic sites. Then the probability that a particular site is occupied by a particular atom is just equal to the atomic fraction of that kind of atom, whatever the nature of the atomic forces. The probability that a particular vector joins two atoms of specified types, however, is dependent on the details of the distribution, and hence on the interatomic forces. Consider a vector  $\mathbf{r}$  linking two atomic positions. There will be many such vectors in a crystal; suppose that in a group  $N(\mathbf{r})$  of them, there are  $N_{AB}(\mathbf{r})$  which connect together an  $A$  atom and a  $B$  atom. Then we can define the probability

$$P_{AB}(\mathbf{r}) = \lim_{N(\mathbf{r}) \rightarrow \infty} \frac{N_{AB}(\mathbf{r})}{N(\mathbf{r})}. \quad (22.1)$$

This is sometimes called a pair density function, by analogy with similar functions used in the descriptions of non-crystalline materials.

Obviously we may define a pair probability function for each possible kind of occupancy of the two sites,  $A-A, A-C, B-B, B-C$ , etc. However, the probabilities so defined are not completely independent, since their values must be consistent with the overall composition of the solid solution. Consider, for example,  $N(\mathbf{r})$  vectors  $\mathbf{r}$  in a binary assembly. They will join together  $2N(\mathbf{r})$  atoms, of which  $2x_A N(\mathbf{r})$  must be of type  $A$ . The vectors linking  $A-A$  atom pairs will contribute  $2N_{AA}(\mathbf{r})$   $A$  atoms to this total, and those linking

$A-B$  pairs will contribute  $N_{AB}(\mathbf{r})$   $A$  atoms. Hence

$$\left. \begin{aligned} 2N_{AA}(\mathbf{r}) + N_{AB}(\mathbf{r}) &= 2x_A N(\mathbf{r}) \\ \text{or } P_{AB}(\mathbf{r}) &= 2\{x_A - P_{AA}(\mathbf{r})\}. \end{aligned} \right\} \quad (22.2)$$

Similarly,

$$P_{AB}(\mathbf{r}) = 2\{x_B - P_{BB}(\mathbf{r})\}.$$

Only one of the three probabilities is independent in this case; we shall take it to be  $P_{AB}(\mathbf{r})$ . There are equivalent relations reducing the number of independent probabilities for ternary and higher assemblies, but we shall only consider binary assemblies in detail. In such an assembly,  $x_A + x_B = 1$ , and it is useful to employ a single composition variable  $x = x_B$ ,  $(1-x) = x_A$ . The composition of the alloy then changes from pure  $A$  to pure  $B$  as  $x$  increases from 0 to 1. Occasionally, however, it is preferable to retain the  $x_A, x_B$  notation in order to show whether complex relations are symmetrical with regard to the components. We shall also need the notation  $N_A = N(1-x)$ ,  $N_B = Nx$  for the numbers of  $A$  and  $B$  atoms in a crystal of  $N$  atoms.

Suppose we have an assembly in which we have values of  $P(\mathbf{r})$  for all possible vectors  $\mathbf{r}$  and all possible combinations of atoms. The question then arises whether or not this is a complete statistical description of the occupancy of the sites. In general, it is not, and we should also write down the probabilities of groups of three, four, five, or more sites being occupied by atom groups of specified types. Such probability factors are introduced in some theoretical treatments, but we shall not have occasion to consider them in detail. The pair probability functions may be deduced from X-ray measurements, as already mentioned, but no experimental method is known for determining triplet or higher probability functions.

A pair probability function must possess the point group symmetry of the Bravais lattice; that is,  $P(\mathbf{r})$  and  $P(\mathbf{r}')$  must be equal if  $\mathbf{r}$  and  $\mathbf{r}'$  are vectors which can be transformed into each other by a rotation or reflection symmetry operation. The behaviour of  $P(\mathbf{r})$  as  $\mathbf{r}$  becomes large falls into one of two categories. In a solid solution which is described as disordered,  $P(\mathbf{r})$  in the limit of large  $\mathbf{r}$  approaches the value it would have for a completely random distribution of atoms over the available sites. This is the situation we shall discuss in most of this chapter. In a structure described as possessing long-range order, the limit of  $P(\mathbf{r})$  as  $\mathbf{r} \rightarrow \infty$  is different from the random value; this is discussed further in Section 26.

In determining the form of an alloy phase diagram, we are interested in the variation of the free energy of a phase with composition. This means that at fixed temperature and pressure, we need to know the heat and entropy of mixing, and various derived functions. The thermodynamic quantities are related to the parameters which specify the distribution of the atoms over the available sites, but any attempt at precise calculation of the energies of different distributions is formidably difficult. Thus, when analytic expressions are required it is necessary to use a simple model. Much of the theoretical work on solid solutions is restricted to binary solutions formed from two metals of the same crystal structure and nearly equal atomic diameters. Under these conditions, the drastic approximations which are used to obtain the required energy expressions are reasonably valid.

We shall assume throughout that the entropy of a solid solution may be separated into configurational and thermal parts. The thermal or vibrational entropy is obtained from the heat motion of the atoms; the configurational entropy results from the randomness of location of the atoms on their sites, and may be defined as the extra entropy possessed by a hypothetical solid solution with all atoms occupying point sites over the corresponding entropy of the pure components. This separation of the entropy is equivalent to a factorization of the partition function.

In the simplest model (the "ideal solution"), the thermal partition function is determined only by the number of atoms of each kind present, and the configurational partition function corresponds to a completely random distribution of the atoms. The formation of an alloy from its constituents then involves no change in thermal entropy, and the entropy of mixing is equal to the configurational entropy of a random arrangement. The assumption of zero thermal entropy of mixing, though unjustified, is often made in more complex treatments, which attempt to calculate the configurational partition function when the arrangement is not random. The model used in most calculations of this type is that of central force nearest neighbour interactions, and the approach is known as the quasi-chemical theory. There are also statistical models in which volume changes and changes in thermal entropy are not neglected, but other simplifying assumptions have to be made in order to obtain useful results. A completely different model focuses attention on the effects of differences in the sizes of the atoms, and calculates the energy term by treating each solute atom as a centre of dilatation in an elastic medium.

These various models are discussed in some detail in the remaining sections of this chapter. None of them is very satisfactory, and at present it seems that experimental measurements of thermodynamic functions are much more accurate than theoretical predictions. In the remainder of this section we examine the important thermodynamic properties of solutions and their relation to the equilibrium diagram.

The thermodynamic properties of a single component are completely determined by the external constraints but have to be measured from an arbitrary zero; that is, a standard reference state must be specified. This is also true of a solution, and it is often convenient to use the pure components under the same temperature and pressure as the reference state. Thus suppose that  $g_A^{z_0}$ ,  $g_B^{z_0}$  are the free energies per atom of the pure components in the  $\alpha$  phase, each measured from an arbitrary zero.† (The  $\alpha$  phase may, of course, be metastable in either or both components.) The free energy per atom of a homogeneous  $\alpha$  solution of composition  $x$  may then be expressed relative to the same zeros as

$$g^\alpha = (1-x)g_A^{z_0} + xg_B^{z_0} + \Delta_m g^\alpha, \quad (22.3)$$

where  $\Delta_m g^\alpha$  is called the free energy of mixing or the free energy of formation of the  $\alpha$  solution, and is measured from the standard state of the unmixed components.

Suppose Fig. 6.1 represents the curve of  $g^\alpha$  against  $x$  at fixed temperature and pressure. An alloy of composition  $x_4$  has a free energy per atom represented by point  $P$  if it exists

† The zeros may be chosen so that at some particular temperature and pressure  $g_A^{z_0} = g_B^{z_0}$ , but this is not then true for other temperatures and pressures.

as a homogeneous  $\alpha$  solid solution, and by point  $Q$  if it is a mixture of pure  $A$  and pure  $B$  (each in the  $\alpha$  form). If the assembly contains two  $\alpha$  solutions of different composition, the  $g^\alpha$  curve gives the free energy per atom in each, and hence the mean free energy per atom. Thus the mean free energy per atom of an alloy of composition  $x_4$  is represented by point  $R$  if it exists as a two-phase mixture of  $\alpha$  solutions of compositions  $x_5$  and  $x_6$ . Clearly, the relative stability of the single solid solution and the phase mixture depends on whether  $R$  is above or below  $P$ . It is readily seen that if the  $g^\alpha$  curve is concave upwards for all  $x$  (i.e. if  $\partial^2 g^\alpha / \partial x^2 > 0$  for all  $x$ ), the single solid solution is always more stable than a phase mixture

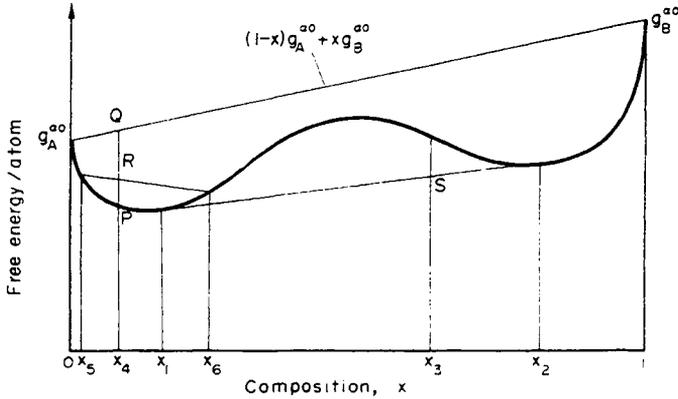


FIG. 6.1. Schematic free energy vs. composition curve for a solid solution.

of two  $\alpha$  solutions. However, if the  $g^\alpha$  curve contains a region of negative  $\partial^2 g^\alpha / \partial x^2$ , as shown in Fig. 6.1, the equilibrium condition will be a phase mixture for certain compositions. For the alloy of composition  $x_3$ , the lowest free energy is represented by the point  $S$ , and is obtained from a mixture of  $\alpha$  phases of compositions  $x_1$  and  $x_2$ . These compositions are those corresponding to the points of contact of the common tangent to the two parts of the  $g^\alpha$  curve. All other alloys in which  $x_1 < x < x_2$  also exist in equilibrium as a mixture of  $\alpha$  phases of compositions  $x_1$  and  $x_2$ , the relative proportions of the phases being fixed by the overall composition.

From this discussion, we see that the limits of solubility  $x_1$  and  $x_2$  are defined by the condition

$$(\partial g^\alpha / \partial x)_{x_1} = (\partial g^\alpha / \partial x)_{x_2} = \{g^\alpha(x_2) - g^\alpha(x_1)\} / (x_2 - x_1). \tag{22.4}$$

Since the  $g^\alpha$  curve is essentially equivalent to the  $\Delta_m g^\alpha$  curve tilted about the straight line  $(1-x)g_A^{\alpha 0} + xg_B^{\alpha 0}$ , this condition may also be written

$$(\partial \Delta_m g^\alpha / \partial x)_{x_1} = (\partial \Delta_m g^\alpha / \partial x)_{x_2}. \tag{22.5}$$

So far as the equilibrium properties are concerned, it is immaterial whether we consider the variation of  $g^\alpha$  or  $\Delta_m g^\alpha$  with  $x$ .

The chemical potential (partial free energy per atom) of  $A$  in the  $\alpha$ -phase is defined by the equation

$$g_A^\alpha = \partial G / \partial N_A = g_A^{\alpha 0} + \partial(N \Delta_m g^\alpha) / \partial N_A. \tag{22.6}$$

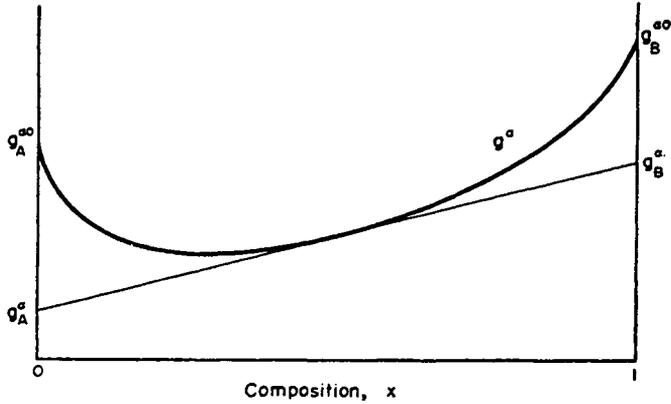


FIG. 6.2. Relation of chemical potentials to free energy vs. composition curve.

The mean free energy per atom is related to the chemical potentials by the equation

and

$$\left. \begin{aligned} g^\alpha &= (1-x)g_A^\alpha + xg_B^\alpha \\ \partial g^\alpha / \partial x &= g_B^\alpha - g_A^\alpha \end{aligned} \right\} \quad (22.7)$$

These relations are illustrated in Fig. 6.2. The tangent to  $g^\alpha$  at the composition  $x$  intercepts the  $x = 0$  ordinate to give the chemical potential  $g_A^\alpha$  and the  $x = 1$  line to give the chemical potential  $g_B^\alpha$ . Clearly, for the common tangent of Fig. 6.1,

$$g_A^\alpha(x_1) = g_A^\beta(x_2), \quad g_B^\alpha(x_1) = g_B^\beta(x_2). \quad (22.8)$$

This is the condition for phase equilibrium in terms of the appropriate intensive thermodynamic quantities (chemical potentials), corresponding to the condition of a minimum value of the extensive quantity  $G$ .

We now consider the possibility of two different phases,  $\alpha$  and  $\beta$ , existing in the assembly. These phases may represent different states of matter, or may be solid solutions of diffe-

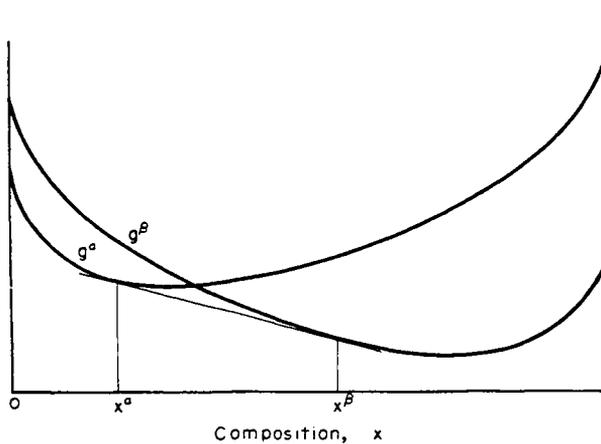


FIG. 6.3. Free energy vs. composition curves for phases  $\alpha$  and  $\beta$ .

rent crystal structure. The free energy of the  $\beta$  phase will be

$$g^\beta = (1-x)g_A^{\beta 0} + xg_B^{\beta 0} + \Delta_m g^\beta, \quad (22.9)$$

where  $g_A^{\beta 0}$ ,  $g_B^{\beta 0}$  are measured from the same zeros as  $g_A^{\alpha 0}$ ,  $g_B^{\alpha 0}$  respectively. Thus  $g_A^{\beta 0} - g_A^{\alpha 0}$  is the free energy per atom required to change pure  $A$  from the  $\alpha$  phase to the  $\beta$  phase. A possible configuration of the free energy curves  $g^\alpha$  and  $g^\beta$  is shown in Fig. 6.3. It follows from the previous discussion that the equilibrium state of the assembly is homogeneous  $\alpha$  phase for  $x < x^\alpha$ , homogeneous  $\beta$  phase for  $x > x^\beta$  and a mixture of  $\alpha$  and  $\beta$  phases for  $x^\alpha < x < x^\beta$ . The solubilities or solvus lines  $x^\alpha$ ,  $x^\beta$  are determined by the relation

$$\left(\frac{\partial g^\alpha}{\partial x}\right)_{x^\alpha} = \left(\frac{\partial g^\beta}{\partial x}\right)_{x^\beta} = \frac{g^\beta(x^\beta) - g^\alpha(x^\alpha)}{x^\beta - x^\alpha}. \quad (22.10)$$

It should be noted that eqn. (22.10) is not equivalent to  $(\partial \Delta_m g^\alpha / \partial x) = (\partial \Delta_m g^\beta / \partial x)$  because of the different reference states used for  $\Delta_m g^\alpha$  and  $\Delta_m g^\beta$ . It is sometimes convenient to rewrite eqns. (22.3) and (22.9) as

$$\left. \begin{aligned} g^\alpha &= (1-x)g_A^{\alpha 0} + xg_B^{\alpha 0} + \Delta_n g^\alpha, \\ g^\beta &= (1-x)g_A^{\beta 0} + xg_B^{\beta 0} + \Delta_n g^\beta, \end{aligned} \right\} \quad (22.11)$$

where  $\Delta_n g^\alpha$ ,  $\Delta_n g^\beta$  are both referred to pure  $A$  and pure  $B$  in their equilibrium states. Equation (22.10) is then equivalent to

$$(\partial \Delta_n g^\alpha / \partial x)_{x^\alpha} = (\partial \Delta_n g^\beta / \partial x)_{x^\beta}. \quad (22.12)$$

Free-energy-composition curves of the type shown in Fig. 6.3 are typical of those expected in a simple eutectic assembly in which the components have different crystal structures  $\alpha$  and  $\beta$ ; those shown in Fig. 6.1 are typical of those from a eutectic assembly in which both metals have the same crystal structure. At temperatures below the eutectic temperature, the liquid free-energy curve lies entirely above the common tangent. Above the eutectic temperature, this curve intersects the common tangent, as shown in Fig. 6.4. Alterna-

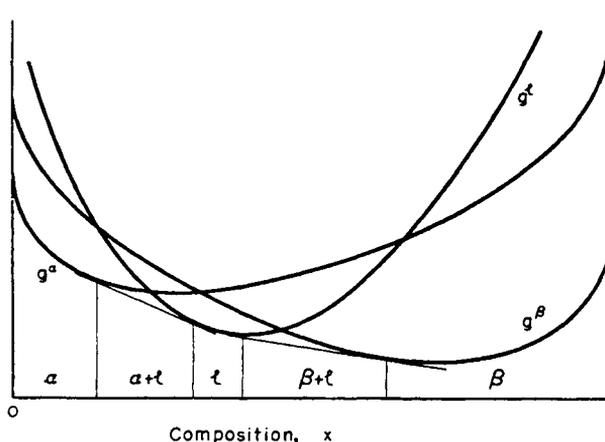


FIG. 6.4. Free energy vs. composition curves for a eutectic reaction.

tively, if the liquid free-energy curve first crosses the  $g^{\alpha}$  curve at  $x = 0$ , or the  $g^{\beta}$  curve at  $x = 1$ , a peritectic reaction will result. The same principles apply when further (intermediate) phases are considered; these may be solid solutions having relatively flat free-energy curves, or compounds of fixed composition, at which the free energy has a very sharp minimum. The relation of free-energy-composition curves to the equilibrium diagram is discussed in many textbooks; see, for example, Cottrell (1955) for an elementary account. It is now possible to store the basic thermodynamic information for a given system in a computer and to use a program which will enable any required information on phase equilibria or derived thermodynamic functions to be obtained. The necessary free-energy functions are generally obtained from a mixture of experimental measurements, empirical expansions, and information derived from measured phase diagrams. Of particular value for work on phase transformations is the ability in this way to obtain reliable estimates of metastable phase relations in otherwise inaccessible regions of the phase diagram; the usual theory of pearlite formation, for example, requires knowledge of the metastable austenite-ferrite and austenite-cementite equilibria below the eutectoid temperature. This kind of work was pioneered by Kaufman (1967, 1969) and has been surveyed by Kaufman and Bernstein (1970) and Hillert (1970).

It is often convenient to refer the thermodynamic properties of a solution to those of an ideal solution. For our present purpose, the simplest definition of an ideal solution is that the heat of mixing is zero, and the entropy of mixing is equal to the configurational entropy of a random arrangement. Consider a simple structure in which  $N$  sites are occupied by  $N_A A$  atoms and  $N_B B$  atoms. The number of distinguishable ways in which the atoms may be arranged on the sites is  $N!/(N_A)!(N_B)!$ , and the configurational entropy of mixing is thus

$$\begin{aligned}\Delta_m S &= k\{\ln N! - \ln N(1-x)! - \ln(Nx)!\} \\ &= -Nk\{(1-x) \ln(1-x) + x \ln x\}\end{aligned}\quad (22.13)$$

and the free energy of mixing per atom is

$$\Delta_m g = kT\{(1-x) \ln(1-x) + x \ln x\}.\quad (22.14)$$

The expression for the entropy of mixing is identical with that obtained previously for lattice defects, which were assumed to be randomly distributed.

In eqns (22.13) and (22.14),  $x$  and  $(1-x)$  are necessarily fractional quantities, less than unity. The free energy of mixing,  $\Delta_m g$ , is thus always negative; it is plotted in Fig. 6.5. The slope of the curve is infinite at  $x = 0$  and  $x = 1$ , and this is sometimes used to justify the statement that pure components cannot form equilibrium phases in an assembly containing two or more components. Very small amounts of a second element dissolved in the lattice will always be distributed randomly, and so give an entropy of mixing corresponding to that of an ideal solution. In a finite crystal, however, solution must be discontinuous, i.e. atom by atom, and there is no reason in principle why the addition of only one atom should not raise the internal energy sufficiently to overcome the associated rise in entropy. This argument is rather academic, and in practice the mixing energy will almost certainly always lead to the presence of a small number of solute atoms in solution at equilibrium, just as

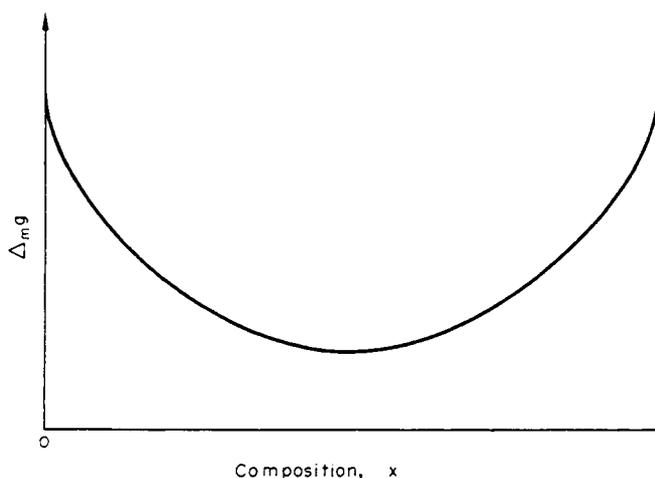


FIG. 6.5. Free energy of mixing vs. composition curve for an ideal solution.

there is always a small number of lattice vacancies. For practical purposes, the solubility may be so small as to be negligible, and we are justified in expecting some solubilities to be too small to be measured by the most sensitive techniques available.

From eqns. (22.6) and (22.4) we may now write expressions for the chemical potentials per atom of an ideal solution. Leaving out the phase identifying superscript,  $\alpha$  or  $\beta$ , since this is common to all quantities, we have

$$\left. \begin{aligned} g_A - g_A^0 &= kT \ln(1-x), \\ g_B - g_B^0 &= kT \ln x. \end{aligned} \right\} \quad (22.15)$$

Guggenheim, and other workers, use a function called the absolute activity,  $\lambda$ , which is related to the chemical potential by the equation  $g_A = kT \ln \lambda_A$ . In terms of these absolute activities, eqn. (22.15) becomes

$$\lambda_A/\lambda_A^0 = 1-x, \quad \lambda_B/\lambda_B^0 = x. \quad (22.16)$$

This equation is useful because it forms the basis of one convenient way of defining the thermodynamic properties of a real solution by means of activity coefficients. For a non-ideal solution, we write

$$\lambda_A/\lambda_A^0 = \gamma_A(1-x), \quad \lambda_B/\lambda_B^0 = \gamma_B x. \quad (22.17)$$

The activity coefficients  $\gamma_A$  and  $\gamma_B$  are functions of composition, and may be determined experimentally for a real solution.

An alternative way of measuring the deviations from "ideal" behaviour is in terms of the "excess" thermodynamic functions. For example, the free energy of mixing may be written

$$\Delta_m g = \Delta_e g + kT\{(1-x) \ln(1-x) + x \ln x\}, \quad (22.18)$$

where  $\Delta_e g$  is called the excess free energy of mixing. The equation

$$\Delta_e g = \Delta_m h - T \Delta_e s \quad (22.19)$$

then defines the (excess) heat of mixing, and the excess entropy of mixing.

The excess quantities just introduced are the mean values per atom. The corresponding partial functions may also be defined in the usual way. For example

$$\Delta_m h = (1-x) \Delta_m h_A + x \Delta_m h_B, \quad (22.20)$$

where  $\Delta_m h_A$ ,  $\Delta_m h_B$  are the partial heats of mixing per  $A$  atom and per  $B$  atom respectively. Thus  $\Delta_m h_A = \partial(N \Delta_m h) / \partial N_A$  is the heat absorbed per atom of  $A$  when a small quantity of  $A$  is added to a very large quantity of solution at constant temperature and pressure; for this reason, it is often called the heat of solution of component  $A$ . Similarly, the excess entropy may be written

$$\Delta_e s = (1-x) \Delta_e s_A + x \Delta_e s_B, \quad (22.21)$$

where  $\Delta_e s_A$ ,  $\Delta_e s_B$  are the excess partial entropies of mixing, or entropies of solution.

In a very dilute solution, the partial excess quantities of the solvent must tend to zero. Thus, there is no heat of solution when  $A$  in the  $\alpha$  phase is dissolved in an  $\alpha$  solution which is already nearly entirely  $A$ , and the entropy of mixing will be represented completely by the configurational term in (22.18). It follows that for a dilute solution

$$\Delta_e g = x(\Delta_m h_B - T \Delta_e s_B) \quad (x \ll 1). \quad (22.22)$$

The  $B$  atoms are sufficiently separated to be considered as non-interacting, and each has associated with it an intrinsic energy, which is the limiting value of  $\Delta_m h_B$  in very dilute solution, and an intrinsic entropy, which is the corresponding limiting value of  $\Delta_e s_B$ . Both of these terms are due essentially to the localized disturbance round each solute atom, and the excess entropy of mixing is entirely due to thermal entropy. The situation in a very dilute solution is exactly analogous to that of a metal containing lattice defects, and the term  $\Delta_m h_B$  corresponds to the intrinsic energy of a defect (e.g.  $\Delta h_{\square}$ ), and  $\Delta_e s_B$  corresponds to the intrinsic thermal entropy of a defect (e.g.  $\Delta s_{\square}$ ).

For very dilute solutions in which  $x_1 \ll 1$ ,  $(1-x_2) \ll 1$ , the equilibrium condition (eqns. (22.4) or (22.5)) becomes

$$(\partial \Delta_m g / \partial x)_{x_1} = (\partial \Delta_m g / \partial x)_{x_2} = 0,$$

so that the variation of the solubility of  $B$  in  $A$  when this solubility is very small is obtained from (22.18) as

$$x_1 = \exp(\Delta_e s_B / k) \exp(-\Delta_m h_B / kT) \quad (22.23)$$

with a similar expression for  $x_2$ .

A similar treatment may be given for the solubility limits in a simple eutectic assembly in which the component metals have different crystal structures, and are very sparingly soluble in each other in the solid. We have now to use the free energies of mixing defined

relative to  $A$  in the  $\alpha$  phase and  $B$  in the  $\beta$  phase (eqn. (22.11)), and the corresponding excess functions which are defined by

$$\Delta_n g = \Delta_f g + kT\{(1-x) \ln(1-x) + x \ln x\}. \quad (22.24)$$

The quantities  $\Delta_n g^\alpha$ ,  $\Delta_f g^\alpha$  differ from  $\Delta_m g^\alpha$  and  $\Delta_e g^\alpha$  respectively by  $x(g_B^{\alpha 0} - g_B^{\beta 0}) = x \Delta g_B^{\beta \alpha}$ , where  $\Delta g_B^{\beta \alpha}$  is the free energy per atom required to change pure  $B$  from the  $\beta$  phase to the  $\alpha$  phase. Similarly,  $\Delta_n g^\beta$  and  $\Delta_f g^\beta$  differ from  $\Delta_m g^\beta$  and  $\Delta_e g^\beta$  respectively by  $(1-x) \Delta g_A^{\alpha \beta}$ .

Writing

$$\Delta_f g^\alpha = x(\Delta_n h_B - T \Delta_f s_B)$$

and using eqn. (22.12), it is readily seen that in very dilute solution the solubility becomes

$$x^\alpha = \exp(\Delta_f s_B/k) \exp(-\Delta_n h_B/kT) \quad (x^\alpha \ll 1), \quad (22.25)$$

which is of the same form as (22.23), and differs only in that

$$\Delta_n h_B = \Delta_m h_B + \Delta h_B^{\beta \alpha}, \quad \Delta_f s_B = \Delta_e s_B + \Delta s_B^{\beta \alpha}. \quad (22.26)$$

It is also possible to obtain an equation of the form (22.23) for the solubility limit when a primary solid solution is in equilibrium with an intermetallic compound; the same procedure is followed, but the free energy of mixing and the corresponding excess functions are defined with reference to the pure metal and the compound as standard state.

Equations (22.23) or (22.25) are only valid for very dilute solutions, when each solute atom has an intrinsic energy and entropy. For moderately dilute solutions ( $x \gtrsim 0.01$ ) Freedman and Nowick (1958) have developed a more general expression by treating eqn. (22.22) as the first term of a Taylor expansion for both  $\Delta_m h$  and  $\Delta_e s$ . For the case of terminal  $\alpha$  and  $\beta$  solutions, they obtain the result

$$\frac{1}{(1-2x^\alpha)} \ln \left\{ \frac{x^\alpha}{1-x^\beta} \right\} = \frac{\Delta_f s_B}{k} - \frac{\Delta_n h_B}{kT}, \quad (22.27)$$

and the equivalent result for two solid solutions of the same structure is clearly obtained by writing  $x^\alpha = x_1$ ,  $x^\beta = x_2$ ,  $\Delta_n h_B = \Delta_m h_B$  and  $\Delta_f s_B = \Delta_e s_B$ . In using this expression to investigate experimental solubility data, the left-hand side is equated to  $\ln x_{\text{corr}}$  thus reducing it to the form of (22.23). A plot of  $\ln x_{\text{corr}}$  against  $1/T$  then enables  $\Delta_n h_B$  and  $\Delta_f s_B$  to be deduced from the slope and the intercept on the  $1/T = 0$  axis of the straight line which should be obtained. In practice, the factor  $1-x^\beta$  may be equated to unity without affecting the results, but the  $1-2x^\alpha$  correction cannot be ignored even in quite dilute solutions, and much better straight lines are obtained from  $\ln x_{\text{corr}}$  than from  $\ln x$ .

An important effect in theories of nucleation (Chapter 10) and growth (Chapter 11) arises from the presence of interfaces in a two-phase assembly. The solubility limits calculated above are strictly valid only for planar interfaces; when interfaces are curved there are additional energy terms which can not be neglected. Consider an assembly in which  $\beta$  particles of surface area  $O$  and surface free energy per unit area  $\sigma$  are in equilibrium with the  $\alpha$  matrix. The energy of the interfaces displaces the equilibrium condition (22.8) or (22.12) and may be considered to contribute an additional thermodynamic potential so

that both the solubility limit  $x^\alpha$  and the equilibrium composition  $x^\beta$  of the  $\beta$  phase vary with the size of the particles. We now denote these quantities as  $x_r^\alpha$  and  $x_r^\beta$  and the equilibrium compositions for infinite planar interfaces by  $x_\infty^\alpha$  and  $x_\infty^\beta$ . Most published treatments of this size effect (Gibbs–Thomson effect) are oversimplified and assume that  $x^\beta$  remains constant.

In a virtual change in which  $dn$  atoms are transferred from the  $\alpha$  phase to the  $\beta$  phase, there is now an additional energy term  $\sigma dO$  due to the increase in surface area of the  $\beta$  particles. As shown in Fig. 6.6, this may be represented as a displacement of  $g^\beta$  to  $g_r^\beta = g^\beta + \sigma(dO/dn)$  and the new equilibrium conditions are given by the points of contact of the tangent common to the curves of  $g^\alpha$  and  $g_r^\beta$  against  $x$ . Let the effective chemical potentials per atom for curved interfaces, defined by the construction of Fig. 6.2 applied

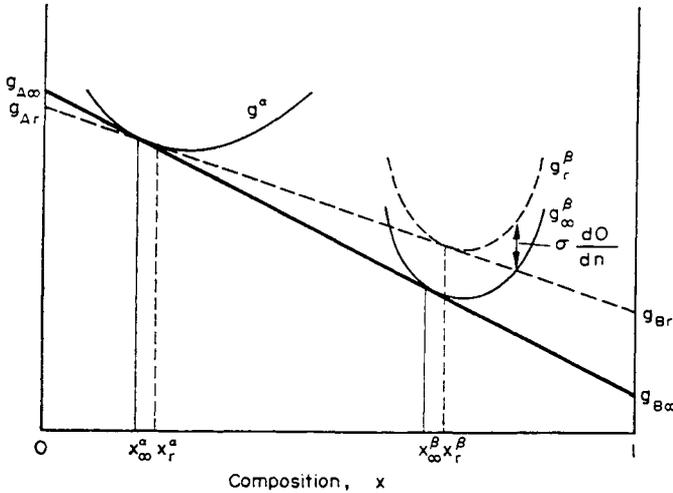


FIG. 6.6. To illustrate the Gibbs–Thomson effect.

to  $g_r^\beta$ , be denoted  $g_{A_r}^\beta$ ,  $g_{B_r}^\beta$ . The new equilibrium conditions are  $g_A^\alpha(x_r^\alpha) = g_{A_r}^\beta(x_r^\beta)$  and  $g_B^\alpha(x_r^\alpha) = g_{B_r}^\beta(x_r^\beta)$  (compare eqn. (22.8)), and we write this more briefly as  $g_{A_r}^\alpha = g_{A_r}^\beta$ , etc. From the geometry of Fig. 6.6, we note that

$$\Delta x_r^\alpha = x_r^\alpha - x_\infty^\alpha \tag{22.28}$$

has the same sign as  $\Delta x_r^\beta$  and is of comparable magnitude, and also that to a good approximation

$$g_{B_r}^\alpha - g_{B_\infty}^\alpha = \{(1 - x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha)\} \sigma(dO/dn) \tag{22.29}$$

In terms of the activity coefficients (eqn. (22.17)), we also have

$$g_{B_r}^\alpha - g_{B_\infty}^\alpha = kT \ln(\gamma_{B_r}^\alpha x_r^\alpha / \gamma_{B_\infty}^\alpha x_\infty^\alpha) \tag{22.30}$$

and for dilute solutions or solutions which obey Henry's law  $\gamma_B^\alpha$  is a constant. Equations (22.29) and (22.30) thus give

$$\frac{\Delta x_r^\alpha}{x_\infty^\alpha} \approx \ln \frac{x_r^\alpha}{x_\infty^\alpha} = \frac{\sigma}{kT} \left( \frac{dO}{dn} \right) \left( \frac{1 - x_\infty^\alpha}{x_\infty^\beta - x_\infty^\alpha} \right). \tag{22.31}$$

In the often used approximation of spherical particles

$$dO/dn = 2v^\beta/r \quad (22.32)$$

and

$$\Delta x_r^\alpha = (2\Gamma/r)x_\infty^\alpha, \quad (22.33)$$

where

$$\Gamma = (\sigma v^\beta/kT)(1-x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha). \quad (22.34)$$

For sufficiently small values of  $x_\infty^\alpha$  (dilute solutions)

$$\Gamma = (\sigma v^\beta/kT)/(x_\infty^\beta - x_\infty^\alpha), \quad (22.35)$$

and this form which has been used by Trivedi (1970) in theories of diffusion-controlled growth (see Chapter 11). A still more extreme assumption is that  $x^\beta = 1$  in which case

$$\Gamma = \sigma v^\beta/kT. \quad (22.36)$$

Equations (22.33) and (22.36) together constitute a frequently quoted (but obviously inaccurate) statement of the Gibbs–Thomson theory. Finally, in the general case where Henry's law does not apply,  $\Gamma$  in eqn. (22.33) is given by

$$\Gamma = (\sigma v^\beta/kT)(1-x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha) \{1 + (d \ln \gamma_B^\alpha / d \ln x)\}. \quad (22.37)$$

Many results obtained with the special assumption of eqn. (22.36) remain valid under more general conditions if  $\Gamma$  from (22.37) is used in place of  $(\sigma v^\beta/kT)$  in the appropriate equations.

If the particles are cylinders (needles) rather than spheres  $dO/dn = v^\beta/r$  and eqn. (22.33) is modified to

$$\Delta x_r^\alpha = (\Gamma/r)x_\infty^\alpha. \quad (22.38)$$

In the general case of equilibrium at a curved interface

$$dO/dn = v^\beta(r_1^{-1} + r_2^{-1}), \quad (22.39)$$

where  $r_1$  and  $r_2$  are the principal radii of curvature. Similarly, for polyhedral particles of equilibrium shape, the surface energy term becomes  $d(O_i\sigma_i)/dn$  where  $O_i$  is a surface in the equilibrium shape at distance  $h_i$  from the centre of the particle so that according to the theory of Section 20,  $\sigma_i/h_i$  is constant. This gives a more general form of (22.33) as

$$\Delta x_r^\alpha = (2\sigma_i/h_i)(\Gamma/\sigma)x_\infty^\alpha. \quad (22.40)$$

Note that since  $\sigma$  is positive  $x_r^\alpha$  is always greater than  $x_\infty^\alpha$ . We may usefully digress here to point out that there is an equivalent relationship between the vapour pressure in equilibrium with a liquid droplet of radius  $r$  and that in equilibrium with a flat liquid surface (Chapter 10, eqn. (46.14)). The curvature of a surface may in fact be regarded as introducing a pressure differential

$$p = \sigma(r_1^{-1} + r_2^{-1}) \quad (22.41)$$

since for a fluid medium the surface tension is numerically equal to the surface free energy  $\sigma$  (Section 35). This concept leads to the same expression for the change in solubility with particle size since the free energy change per atom is now  $pv^\beta$ .

Equations (22.31), (22.33), (22.37), and (22.40) and the equivalent vapour pressure eqn. (46.14) are variously known as the Gibbs–Thomson, Thomson–Helmholtz, or Thomson–Freundlich equation (Gibbs 1878; Thomson 1888; Freundlich 1926) or sometimes simply as the effects of “capillarity”, which is the term used by Gibbs.<sup>†</sup> The equations are often written in terms of concentrations  $c_{\infty}^{\alpha}$ ,  $c_r^{\alpha}$ , etc., instead of atomic fractions  $x_{\infty}^{\alpha}$ ,  $x_r^{\alpha}$ , etc, but this is strictly valid only when there is negligible volume change on transformation. The above treatment is implicitly based on this assumption since no allowance is made for strain energies due to transformation (see below).

We conclude this section by examining the thermodynamics of inhomogeneous phases and the criteria for distinguishing between a metastable phase and an unstable phase. A solid solution of fluctuating composition may be assigned a free energy which contains three different terms. The first contribution is the sum of the free energies which the individual atoms would have if they were present in homogeneous solutions of the same compositions as those of their local environments. Each term in the sum is thus represented by a point on the free-energy curve of the homogeneous solution. In the particular case that there are regions of only two different compositions, the change in free energy per atom in comparison with the homogeneous solution is given graphically by the construction shown in Fig. 6.1, and the sum over  $N$  atoms is just a generalization of this procedure. The second term arises because the free energy of any small volume element of given composition will be changed if it is surrounded by material of different composition. When the transition is sharp, the extra energy appears macroscopically as the surface free energy we have discussed above; Cahn and Hilliard (1958) showed how to express the analogous energy which arises from a continually fluctuating composition. The third term was mentioned briefly above; it is the elastic coherency strain energy which is due to a variation of lattice parameter with composition. If all regions of the inhomogeneous phase are coherent with each other, the assembly will be in a self-stressed state, the elastic energy of which must be considered part of the overall free energy.

We consider first the chemical or volume free energy. For *small* variations about the mean composition, the generalized construction of Fig. 6.1 will always leave  $R$  above  $P$  in those composition ranges where  $(\partial^2 g/\partial x^2) > 0$ , and this includes some compositions between  $x_1$  and  $x_2$ . However, there is a net decrease in free energy, even for very small fluctuations, if  $(\partial^2 g/\partial x^2) < 0$ . The curve

$$(\partial^2 g/\partial x^2) = (\partial^2 \Delta_m g/\partial x^2) = 0 \quad (22.42)$$

was called by Gibbs the limit of metastability, since inside this curve a homogeneous solution should decompose spontaneously at a rate limited only by the rate of atomic migration. The curve is now usually referred to as the spinodal, or more strictly as the chemical spinodal. As we shall describe in later chapters, the decomposition of a supersaturated solid solution inside the spinodal does not require nucleation in the classical sense and occurs by a process known as “uphill” or “negative” diffusion. Note that there is no spinodal in free-energy diagrams such as those shown in Figs. 6.2–6.5.

<sup>†</sup> An interesting survey of the early history of this subject is given by Swan and Urquhart (1927).

Cahn and Hilliard used a multivariable Taylor expansion to express the free energy of an inhomogeneous solution in terms of  $g^a$  and of the various spatial derivatives of composition  $x$  (or of concentration  $c$ ). The first term in this expansion is the sum already considered and for centrosymmetric crystals the energy per atom in the next most important term may be written as  $Kv(\nabla x)^2 = Kv^3(\nabla c)^2$ , where  $v$  is the atomic volume and  $K$  is a second rank tensor and is thus isotropic for cubic crystals. This "gradient energy" is a continuum analogue of the surface free energy of a discontinuous interface; the Cahn and Hilliard expansion is, of course, invalid if the gradient  $\nabla c$  becomes large. There has been some dispute over the validity of the symmetry argument for the absence of a term in  $|\nabla c|$  (see Tiller *et al.* 1970; Cahn and Hilliard, 1971), but we believe the Cahn and Hilliard treatment is correct.

An exact theory of the coherency energy is difficult when allowance is made for elastic anisotropy (see Cahn 1969), and we mention here only the isotropic result. The elastic energy depends on  $\varepsilon = d \ln a/dx$ , where  $a$  is the lattice parameter, and on an elastic modulus  $Y' = Y/(1-\nu)$ , where  $Y$  is Young's modulus and  $\nu$  is Poisson's ratio. Cahn (1961) showed that if the actual composition is  $x$  and the mean composition is  $x_0$  the coherency energy per atom may be written as  $\varepsilon^2 Y' v (x - x_0)^2$ . This changes both the solubility limits and the limits of metastability. In place of eqn. (22.4), the coherent miscibility gap is defined by  $x_{1c}$  and  $x_{2c}$ , where

$$\left( \frac{\partial g^a}{\partial x} \right)_{x_{1c}} = \frac{g(x_{2c}) - g(x_{1c})}{x_{2c} - x_{1c}} + 2v\varepsilon^2 Y' (x_{2c} - x_{1c}) \quad (22.43)$$

with a similar equation for  $x_{2c}$  and

$$x_{1c} > x_1, \quad x_{2c} < x_2. \quad (22.44)$$

Similarly, the coherent spinodal which represents the true limit of metastability is defined by

$$(\partial^2 g / \partial x^2) + 2\varepsilon^2 Y' v = 0, \quad (22.45)$$

and lies entirely inside the chemical spinodal. The increase in solubility  $x_{1c} - x_1$  at fixed temperature, or equivalently the depression of the temperature of the coherent curves below the chemical curve, depends sensitively on the magnitude of  $\varepsilon$ ; the supercooling from the chemical to the coherent spinodal was estimated by Cahn as 40°C for aluminium-zinc alloys and 200°C for gold-nickel alloys.

Note that although the coherency strain energy is a function of the mean composition  $x_0$  of the alloy, the positions of the coherent phase boundary and spinodal are not dependent on  $x_0$ . However, directional effects appear when anisotropic elastic theory is used; for example, the solid solution may be unstable to fluctuations in composition in some directions in a crystal but metastable for other directions (see Part II, Chapter 18).

Equation (22.44) shows that the coherent spinodal only exists when there is a negative region of  $(\partial^2 g / \partial x^2)$ , i.e. when there is also a chemical spinodal. However, precipitates of one phase in another which involve chemical free energy curves of the type shown in Fig. 6.3 may sometimes be obtained in either coherent or incoherent forms, and there is then in principle both a coherent solvus and an equilibrium solvus. Intermediate or metastable phases which are coherent with the matrix are also often encountered in low-temperature

precipitation reactions. Such phases frequently have a lower symmetry than the equilibrium phase and it may occasionally be illuminating to regard them as coherent versions of the equilibrium precipitate.

### 23. THE NEAREST NEIGHBOUR MODEL : REGULAR SOLUTIONS

The forces between the atoms in a solid solution, as in a pure metal, are mainly short-range in character, and the most used model is that in which only nearest neighbour interactions are considered significant. The energy of the crystal is then the sum of the pair interaction energies of the  $A-A$ ,  $B-B$ , and  $A-B$  contacts. This is often referred to as the Ising model, although Ising actually treated only the magnetic analogue of this situation; the first application to alloys was made by Bethe.

The limitations of the nearest neighbour model in providing a satisfactory description of the metallic state have already been emphasized in the previous chapter. When applied to solid solutions, an additional difficulty is at once apparent, since the theory predicts that all properties of the solid solution are symmetrical about the 50 atomic% composition, and this is rarely true in practice. This limitation may be avoided by modifying the theory, but at least two parameters are then required to specify the energy of the assembly. A major attraction of the simple central force model (to the mathematician, at least!) is that all properties of the solid solution at fixed temperature, and of the equilibrium diagram, can be described in terms of a single arbitrary parameter.

Let us consider a binary alloy containing  $A$  and  $B$  atoms. In place of the characteristic nearest neighbour interaction energy,  $-2\bar{\mathcal{E}}$ , we now have the energies  $-2\mathcal{E}_{AA}$ ,  $-2\mathcal{E}_{BB}$ , and  $-2\mathcal{E}_{AB}$  representing the binding energies of two  $A$  atoms, two  $B$  atoms and an  $A$  atom and a  $B$  atom respectively. The model assumes that  $\mathcal{E}_{AA}$ ,  $\mathcal{E}_{BB}$ , and  $\mathcal{E}_{AB}$  are all independent of the surrounding configuration, and this is probably its most serious limitation.

The three potential energy terms are independent of each other, but in the nearest neighbour model, the properties of the solid solution depend only on the combination

$$\bar{\mathcal{E}} = \mathcal{E}_{AA} + \mathcal{E}_{BB} - 2\mathcal{E}_{AB}, \quad (23.1)$$

The nature of the quantity  $\bar{\mathcal{E}}$  may be seen by a hypothetical process in which we interchange an  $A$  atom and a  $B$  atom on any two sites in the crystal. If the number of  $A-A$  contacts increases or decreases by  $X$ , it follows that whatever the arrangement around the two sites, the number of  $B-B$  contacts also changes by  $X$ , and the number of  $A-B$  contacts by  $-2X$ . The change in energy is thus  $-X(\mathcal{E}_{AA} + \mathcal{E}_{BB} - 2\mathcal{E}_{AB})$ .

In an ideal solid solution, the internal energy of the crystal is independent of the atomic arrangement. This condition is satisfied if  $\mathcal{E}_{AA} = \mathcal{E}_{BB} = \mathcal{E}_{AB}$ , and (less restrictively) if  $\bar{\mathcal{E}} = 0$ . The condition for the formation of an ideal solid solution with this model is thus that the force between unlike atoms is equal to the average of the forces between two like atoms. When longer-range forces are considered, the corresponding conditions are obvious; in any interchange of  $A$  and  $B$  atoms within the interior of the crystal, the change in energy must vanish. One way of including longer-range forces is to use a pairwise interaction model

with a series of interchange potentials  $\mathcal{E}_1, \mathcal{E}_2$ , etc., defined by

$$\mathcal{E}_i = \mathcal{E}_{AA,i} + \mathcal{E}_{BB,i} - 2\mathcal{E}_{AB,i} \quad (23.1a)$$

where each potential is defined at a separation corresponding to the *i*th nearest neighbour distance. Note that although, according to this model, there are three interaction potentials for each distance of separation, it is impossible to imagine an experiment which would enable the individual potential to be measured separately, and only the interchange potentials  $\mathcal{E}_i$  have operational significance. In any interchange of an *A* atom and a *B* atom, the change in energy is now given by the sum  $-X_i\mathcal{E}_i$ , where  $X_i$  is the increase in the number of *i*th neighbour *A-A* or *B-B* bonds, and  $-2X_i$  is the corresponding increase in *A-B* bonds.

For many solid solutions of close-packed structures, the limitations of the nearest neighbour model lie more in the assumption of central forces and the independence of the  $\mathcal{E}$  terms on the environment than in the neglect of second and third nearest neighbour interactions. Thus Guggenheim (1952) showed that if the interaction energy varies as  $r^{-6}$ , the effect of second nearest neighbours on measurable thermodynamic quantities is negligible. The use of such interactions makes equations unwieldy and only in special circumstances does it lead to any essential improvement in the physical description. We shall therefore confine ourselves to nearest neighbour forces in this section, although there are then obvious difficulties in applying the theory to b.c.c. structures which are not mechanically stable under nearest neighbour forces. It will be necessary to consider higher neighbour interactions in connection with order-disorder changes (Section 26), and some comments will also be made there on their influence in systems exhibiting phase segregation.

Non-ideal solutions, which have  $\mathcal{E} \neq 0$ , may be classified qualitatively by the sign of  $\mathcal{E}$ . If  $\mathcal{E}_{AA} + \mathcal{E}_{BB} < 2\mathcal{E}_{AB}$ , the attractive forces between like atoms are weaker than those between unlike atoms, and there will be a tendency for each atom to surround itself with as many atoms of the opposite kind as possible. This ordering tendency may produce a superlattice at low temperatures; at high temperatures, it is opposed by the thermal energy, which always tends to produce a random arrangement of high entropy. For  $\mathcal{E}$  positive, the opposite result is valid, and at low temperatures the solid solution tends to segregate into *A*-rich and *B*-rich regions.

When the solution is not ideal, theoretical expressions can most readily be obtained on the assumption that it is "regular". This term was first introduced by Hildebrand to describe a class of solutions having physical properties which vary with composition in a regular manner; the definition of the regular solution has, however, varied considerably amongst different workers. The earliest approach was to define a regular solution as a solution in which the configurational entropy of mixing is still given by eqn. (22.13), even though the heat of mixing is not zero. The atomic arrangement is thus considered to be effectively random, although interchanges of atoms lead to changes in the internal energy. This assumption is obviously roughly justified if the magnitude of  $\mathcal{E}$  is small. In this sense, a regular solution is one which deviates only slightly from ideal conditions.

Guggenheim (1952) uses a less restricted definition of regular solution, and regards the above definition as a crude or "zeroth order" approximation. This approach has some ad-

vantage in emphasizing the relation between the simple theory and the higher approximations of the quasi-chemical theory.

For a simple structure in which each site has a common coordination number  $z$ , there will be  $\frac{1}{2}Nz$  nearest neighbour bonds in a crystal of  $N$  atoms, provided  $N$  is sufficiently large for surface effects to be negligible. We write the total number of nearest neighbour  $A-B$  contacts as  $zN_{AB}$  (i.e.  $zN_{AB}$  is the value of  $N_{AB}(\mathbf{r})$  for the whole crystal when  $\mathbf{r}$  is a nearest neighbour vector).<sup>†</sup> Similarly,  $zN_{AA}$  and  $zN_{BB}$  are the total numbers of  $A-A$  and  $B-B$  contacts in the crystal. By considering the neighbours of the  $A$  atoms and the  $B$  atoms separately, we find

$$\left. \begin{aligned} \text{Number of } A-A \text{ contacts} &= zN_{AA} = \frac{1}{2}z(N_A - N_{AB}), \\ \text{Number of } B-B \text{ contacts} &= zN_{BB} = \frac{1}{2}z(N_B - N_{AB}), \\ \text{Number of } A-B \text{ contacts} &= zN_{AB}. \end{aligned} \right\} \quad (23.2)$$

For a random distribution,  $P_{AA}(\mathbf{r}) = (1-x)^2$ ,  $P_{BB}(\mathbf{r}) = x^2$ , and  $P_{AB}(\mathbf{r}) = 2x(1-x)$  for all possible interatomic vectors  $\mathbf{r}$ . In particular, the numbers of nearest neighbour pairs of the types  $A-A$ ,  $B-B$ ,  $A-B$  will be  $\frac{1}{2}Nz(1-x)^2$ ,  $\frac{1}{2}Nzx^2$ , and  $Nzx(1-x)$ . The potential energy of the crystal is thus

$$\left. \begin{aligned} U &= -Nz\{(1-x)^2\mathcal{E}_{AA} + x^2\mathcal{E}_{BB} + 2x(1-x)\mathcal{E}_{AB}\} \\ &= -Nz\{(1-x)\mathcal{E}_{AA} + x\mathcal{E}_{BB} - x(1-x)\mathcal{E}\}. \end{aligned} \right\} \quad (23.3)$$

The value of  $U$  for the components is  $-Nz(1-x)\mathcal{E}_{AA} - Nz x\mathcal{E}_{BB}$ , so that we have

$$\Delta_m H = \Delta_m U = Nz x(1-x)\mathcal{E} \quad (23.4)$$

as the heat or energy of mixing.<sup>‡</sup> This curve has a simple  $U$ -shape, with a maximum or minimum at  $x = \frac{1}{2}$ , according to the sign of  $\mathcal{E}$ . The configurational entropy of mixing is given by eqn. (22.13), since the atomic arrangement is assumed to be random. If we assume for simplicity that the thermal entropy of mixing is zero, we obtain for the free energy of mixing per atom

$$\Delta_m g = zx(1-x)\mathcal{E} + kT\{(1-x)\ln(1-x) + x\ln x\}. \quad (23.5)$$

The form of the curve of  $\Delta_m g$  against composition thus depends on the sign of  $\mathcal{E}$ . The relation is shown in Fig. 6.7 for  $\mathcal{E}$  negative (tendency to form superlattices),  $\mathcal{E}$  zero (ideal solutions) and  $\mathcal{E}$  positive (tendency for phase segregation). The latter case corresponds to Fig. 6.1, and the compositions of the two phases in equilibrium are given by eqn. (22.5). In the present model, eqn. (23.5) is symmetrical about  $x = \frac{1}{2}$ , and the common tangent to the two minima in the curve has zero slope, so that

$$(\partial\Delta_m g/\partial x)_{x_1} = (\partial\Delta_m g/\partial x)_{x_2} = 0. \quad (23.6)$$

<sup>†</sup> It is more convenient to write the number of  $A-B$  contacts as  $zN_{AB}$  than as  $N_{AB}$ ; this means that  $P_{AB}(\mathbf{r}) = 2N_{AB}/N$  for nearest neighbour contacts, and simplifies the algebra of the quasi-chemical theory.

<sup>‡</sup> In condensed phases, we need not distinguish between  $\Delta U$  and  $\Delta H$ , or between  $\Delta F$  and  $\Delta G$ .

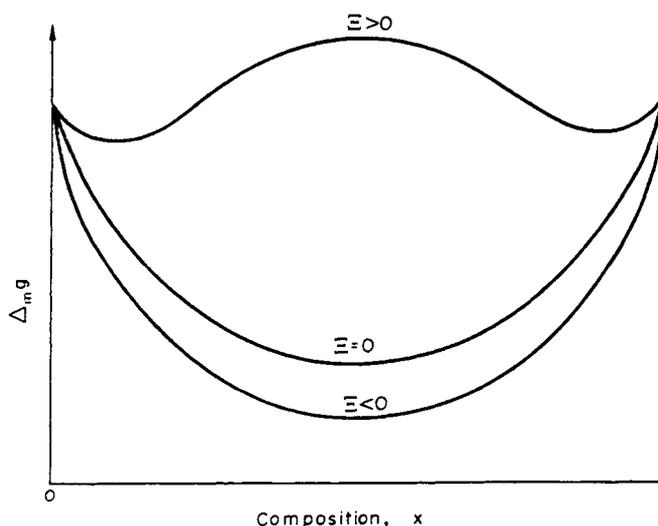


FIG. 6.7. Free energy of mixing vs. composition curves according to (23.5).

From eqns. (23.5) and (23.6),

$$\left(\frac{1}{2x-1}\right) \ln\left(\frac{x}{1-x}\right) = \frac{z\Xi}{kT}. \quad (23.7)$$

At temperatures where the curve of  $\Delta_m g$  has the form of Fig. 6.1, there are three solutions to (23.7). The solution at  $x = \frac{1}{2}$  gives a maximum value to  $\Delta_m g$ ; the other two solutions,  $x_1$  and  $x_2$ , give minima, and are symmetrical about  $x = \frac{1}{2}$ . For dilute solutions ( $x \ll 1$ ), we write as an approximation  $(1-x) \approx (1-2x) \approx 1$ , and the variation of the solubility limit  $x_1$  is then

$$x_1 = \exp-(z\Xi/kT), \quad (23.8)$$

so that the solubility approximates to an exponential curve if the solubility is small, i.e. if  $z\Xi/kT$  is large.

The similarity of eqn. (23.7) to Freedman and Nowick's expression eqn. (22.27) should be noted, but the underlying physical assumptions are different. Equation (22.27) applies to any solubility limit if the solubility is small, and includes the effects of thermal entropy. Equation (23.7) applies only to the artificial regular solution model, but within this limitation it is valid for all values of  $x$ . The limiting forms of the two equations in very dilute solutions (eqns. (22.23) and (23.8)) differ insofar as a thermal entropy contribution from each solute atom is present in a real solution, but not in the model.

Equation (23.7) is sometimes written in the form

$$(2/(2x-1)) \tanh^{-1}(2x-1) = z\Xi/kT. \quad (23.9)$$

As the temperature is raised, the two roots  $x_1$  and  $x_2$  both approach the value  $x = \frac{1}{2}$  until, finally, they coincide at this value, and the curve of  $\Delta_m g$  assumes a simple U-form. At all

temperatures above this critical temperature, there is a continuous solid solution from pure  $A$  to pure  $B$ . The  $\Delta_m g$  curve has a maximum at  $x = \frac{1}{2}$  for  $T < T_c$ , and a minimum at  $x = \frac{1}{2}$  for  $T > T_c$ , so that  $T_c$  is given by the condition  $(\partial^2 \Delta_m g / \partial x^2)_{x=1/2} = 0$ . Using eqn. (23.5),

$$T_c = z\bar{\epsilon}/2k. \quad (23.10)$$

The complete solubility curve for the hypothetical assembly we are considering is plotted in Fig. 6.8, the temperatures being measured in units of  $kT/z\bar{\epsilon}$ . The assumptions lead to an equilibrium diagram in which there is a solubility gap below  $T_c$ , the boundaries of the gap being symmetrical about the equi-atomic composition. The solubilities are zero at 0 K,

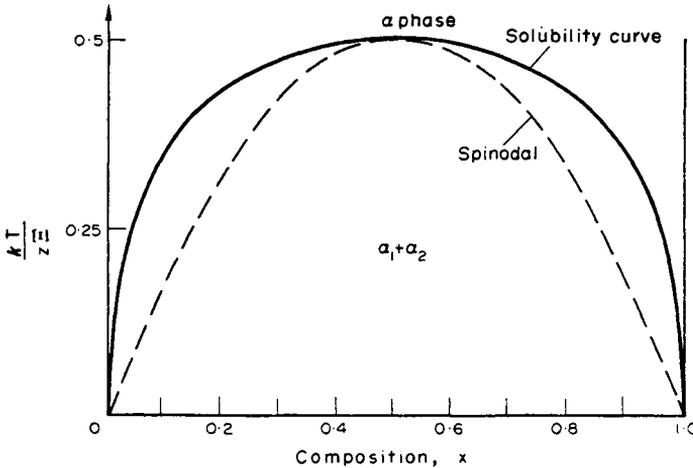


FIG. 6.8. The solubility limits and spinodal curve given by eqns. (23.9) and (23.11).

since the formation of a solid solution raises the internal energy of the assembly above that of a phase mixture of the pure components. As already noted, the equilibrium diagram of Fig. 6.8 is simply related to that of a eutectic assembly in which two metals of the same crystal structure have a limited mutual solubility in the solid state. The eutectic reaction occurs if the liquid phase becomes stable before the solubility gap is closed.

In addition to the boundaries of the two-phase region we are also interested in the chemical spinodal [eqn. (22.42)] which is given by

$$x(1-x) = kT/2z\bar{\epsilon} \quad (23.11)$$

in a regular solution. At all temperatures below  $T_c = z\bar{\epsilon}/2k$ , the roots of this equation lie inside the roots  $x_1$  and  $x_2$  of eqn. (23.9). The two curves, which touch at  $T = 0$  K and  $T = T_c$ , are shown in Fig. 6.8.

The chemical potential of  $A$  in the regular solution is given by

$$g_A - g_A^0 = z\bar{\epsilon}x^2 + kT \ln(1-x)$$

from eqn. (23.5). The activity coefficient is thus

$$\gamma_A = \exp(z\bar{\mathcal{E}}x^2/kT), \quad (23.12)$$

and we have, similarly,

$$\gamma_B = \exp\{z\bar{\mathcal{E}}(1-x)^2/kT\}.$$

#### 24. THE QUASI-CHEMICAL THEORY: OTHER STATISTICAL THEORIES

We now consider the possibility of removing some of the very restrictive assumptions used in the zeroth order approximation of the treatment of regular solutions. The simple theory above is incorrect because of the neglect of thermal entropy factors, and in the assumption of random atomic arrangement. Both these factors may be important, but it is difficult to devise a model which will include them together in a satisfactory manner. In the quasi-chemical theory, which we shall describe first, emphasis is placed on improving the calculation of the configurational entropy, and the properties of the solution are again expressed in terms of the single parameter  $\bar{\mathcal{E}}$ . In the usual application,  $\bar{\mathcal{E}}$  is temperature independent, as assumed above, and the only contribution to  $\Delta_m s$  comes from configurational terms. Thermal entropy terms can be formally included if  $\bar{\mathcal{E}}$  is allowed to vary with temperature.

In a real solution, it is physically obvious that the atom distribution will be random only at very high temperatures, and the configurational entropy of mixing must therefore be a function of temperature. Since a random distribution gives the maximum number of distinguishable arrangements which the assembly can possess, a more accurate calculation of the configurational entropy term must lead to a negative excess entropy of mixing if the thermal entropy of mixing is ignored. Consider the assembly in a particular macroscopic configuration which we may specify by the number of nearest neighbour  $A-B$  pairs  $zN_{AB}$ . To find the equilibrium state, we have to calculate the number of microscopic states leading to this configuration, and the energy of the configuration. We can then take the sum over all values of  $N_{AB}$ , and thus obtain the partition function.

The total energy of the assembly is  $U = -z(N_A\bar{\mathcal{E}}_{AA} + N_B\bar{\mathcal{E}}_{BB} - N_{AB}\bar{\mathcal{E}})$  (using eqn. (23.2)), and the heat of mixing is thus

$$\Delta_m H = \Delta_m U = z\overline{N_{AB}}\bar{\mathcal{E}}, \quad (24.1)$$

where  $\overline{N_{AB}}$  is the equilibrium value of  $N_{AB}$ . Denoting the total number of arrangements for given  $N_{AB}$  by  $\Omega(N_A, N_B, N_{AB})$ , we have the partition function

$$Q = \sum_{N_{AB}} \Omega(N_A, N_B, N_{AB}) \exp(-U/kT). \quad (24.2)$$

This partition function has not been evaluated exactly for a three-dimensional lattice. An exact solution in two dimensions was obtained by Onsager (1944) using a complex mathematical method, and his results have been extended by other workers using simpler methods.

In three dimensions, the value of  $\Omega(N_A, N_B, N_{AB})$  may be evaluated if we assume as an approximation that the various kinds of pair do not interfere with each other, and may be

treated as separate entities. We note first of all that

$$\Sigma\Omega(N_A, N_B, N_{AB}) = N!/(N_A)!(N_B)! \quad (24.3)$$

since this is merely the expression for the total number of arrangements of the assembly. We are going to be concerned with the logarithm of the function  $\Omega$ , and we may then replace the sum on the left of this expression by the maximum term; this procedure is very commonly used in statistical thermodynamics. Suppose the maximum value of  $\Omega(N_A, N_B, N_{AB})$  is obtained for some particular value  $N_{AB}^0$ . We then have

$$\Omega(N_A, N_B, N_{AB}^0) = N!/(N_A)!(N_B)!$$

The assumption that the total number of arrangements for given  $N_{AB}$  is obtained by treating the various pairs as independent entities requires that  $\Omega(N_A, N_B, N_{AB})$  is proportional to

$$\frac{(\frac{1}{2}zN)!}{\{\frac{1}{2}z(N_A - N_{AB})\}! \{\frac{1}{2}z(N_B - N_{AB})\}! (\frac{1}{2}zN_{AB})! (\frac{1}{2}zN_{BA})!}$$

in which we have allowed for the orientation of the sites, which effectively distinguishes an  $A-B$  pair from a  $B-A$  pair. The number of arrangements is only proportional to the above expression, not equal to it, since the total number of arrangements must satisfy (24.3). We can achieve this by introducing a normalization factor, so that from (24.3) we find

$$\Omega(N_A, N_B, N_{AB}) = \frac{N!}{N_A! N_B!} \frac{\{\frac{1}{2}z(N_A - N_{AB}^0)\}! \{\frac{1}{2}z(N_B - N_{AB}^0)\}! \{\frac{1}{2}zN_{AB}^0\}! \{\frac{1}{2}zN_{AB}^0\}!}{\{\frac{1}{2}z(N_A - N_{AB})\}! \{\frac{1}{2}z(N_B - N_{AB})\}! \{\frac{1}{2}zN_{AB}\}! \{\frac{1}{2}zN_{AB}\}!} \quad (24.4)$$

Also, by differentiating expression (24.4) with respect to  $N_{AB}$ , we obtain the maximum value of  $N_{AB}$  which is given by

$$0 = \partial\Omega(N_A, N_B, N_{AB})/\partial N_{AB} = \partial \ln \Omega(N_A, N_B, N_{AB})/\partial N_{AB}.$$

On taking logarithms and using Stirling's theorem, this reduces to

$$N_{AB}^0 = N_A N_B / N = Nx(1-x), \quad (24.5)$$

which is the value of  $N_{AB}$  for a completely random arrangement, as anticipated.

We can now rewrite eqn. (24.2) as

$$Q = \Sigma\Omega(N_A, N_B, N_{AB}) \exp\{-z(N_A\bar{\epsilon}_{AA} + N_B\bar{\epsilon}_{BB} + N_{AB}\bar{\epsilon})/kT\}.$$

and we now have an explicit value for  $\Omega$ . Once again we replace the sum by its maximum term; this corresponds to  $N_{AB}$  having its equilibrium value  $\bar{N}_{AB}$ . Thus

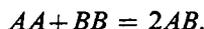
$$Q = \Omega(N_A, N_B, \bar{N}_{AB}) \exp\{-z(N_A\bar{\epsilon}_{AA} + N_B\bar{\epsilon}_{BB} + \bar{N}_{AB}\bar{\epsilon})/kT\} \quad (24.6)$$

and  $\partial \ln Q/\partial N_{AB} = 0$  for  $N_{AB} = \bar{N}_{AB}$ . We have, therefore, to substitute eqn. (24.4) into

(24.6) and differentiate in order to find  $\overline{N_{AB}}$ . Proceeding as we did for  $N_{AB}^0$ , we find that

$$(\overline{N_{AB}})^2 = (N_A - \overline{N_{AB}})(N_B - \overline{N_{AB}}) \exp(-2\Xi/kT). \quad (24.7)$$

This is the basic formula of the present approximation, and was originally derived by Guggenheim by a different method. He treated the interchange of an  $A$  atom and a  $B$  atom as a chemical process in which  $A-A$  bonds and  $B-B$  bonds "react" to form  $A-B$  bonds, and vice versa:



If the law of mass action is applied to this symbolic equation, we obtain the result (24.7). For this reason, the method is often called the quasi-chemical theory. Guggenheim has also shown it is equivalent to the method which Bethe developed for order-disorder reactions (see p. 221).

Equation (24.7) is a rather unpleasant quadratic. The algebra is simplified by writing the solution in the form

$$\overline{N_{AB}} = 2N_{AB}^0/(\beta+1) = 2Nx(1-x)/(\beta+1). \quad (24.8)$$

Substituting this into (24.7) shows that  $\beta$  is the positive root of the equation

$$\beta^2 - (1-2x) = 4x(1-x) \exp(2\Xi/kT). \quad (24.9)$$

Equation (24.8) shows that  $\beta = 1$  in the zeroth approximation, and  $\beta > 1$  in the present treatment. It follows from (24.1) and (24.8) that the heat of mixing per atom is

$$\Delta_m h = \Delta_m u = 2zx(1-x)\Xi/(\beta+1). \quad (24.10)$$

We can now deduce the other thermodynamic properties of the assembly, using the relation

$$G = F = -kT \ln Q = -kT \ln \Omega(N_A, N_B, \overline{N_{AB}}) + U.$$

From eqn. (24.6) it follows that  $\partial F/\partial \overline{N_{AB}} = 0$ , and we already know that  $\partial \ln \Omega/\partial N_{AB}^0 = 0$ .

We may also write:

$$\begin{aligned} \Delta_m g &= x_A(g_A - g_A^0) + x_B(g_B - g_B^0) = kT \left[ (1-x) \ln(1-x) + x \ln x \right. \\ &\quad \left. + \frac{1}{2} z \left\{ x_A \ln \left( \frac{N_A - \overline{N_{AB}}}{N_A - N_{AB}^0} \right) + x_B \ln \left( \frac{N_B - \overline{N_{AB}}}{N_B - N_{AB}^0} \right) \right\} \right] \\ &\quad + 2zx(1-x)\Xi/(\beta+1). \end{aligned} \quad (24.11)$$

This equation is simple and symmetrical.† The quantities  $\overline{N_{AB}}$  and  $N_{AB}^0$  are defined in eqn. (24.7) and (24.5); as already noted, they have simple physical interpretations.  $N_{AB}^0$  is the number of  $A-B$  bonds for a random mixture, and  $\overline{N_{AB}}$  is the mean number of  $A-B$  bonds

† The internal energy  $U$  was inadvertently left out of eqn. 24.11 (24.13 in previous printings). I am grateful to Professor H.K.D.H. Bhadeshia for pointing this out to me.

for the actual solution. If we wish to eliminate the numbers  $N_A$ ,  $N_B$ ,  $\overline{N_{AB}}$ , we may write the equation in the form

$$\begin{aligned} \Delta_m g = kT & \left[ (1-x) \ln(1-x) + x \ln x \right. \\ & \left. + \frac{1}{2} z \left\{ (1-x) \ln \frac{\beta+1-2x}{(1-x)(\beta+1)} + x \ln \frac{\beta-1+2x}{x(\beta+1)} \right\} \right] \\ & + 2zx(1-x)\mathcal{E}/(\beta+1). \end{aligned} \quad (24.12)$$

From eqns. (24.10) and (24.12) an expression for the excess entropy of mixing may be obtained.

To find the equilibrium state, we must equate  $\partial\Delta_m g/\partial x = 0$ . As in the zeroth approximation, we find  $x = \frac{1}{2}$  is always a solution, but that this gives a maximum value to  $\Delta_m g$  below some critical temperature, and a minimum value above this temperature. Below the critical temperature, there are two other solutions symmetrically disposed about  $x = \frac{1}{2}$ .

The differentiation of eqn. (24.12) involves some rather lengthy algebra because  $\beta$  is a function of  $x$  (eqn. (24.9)). Equating the first differential coefficient to zero, we obtain

$$\ln \frac{x}{1-x} + \frac{z}{2} \ln \frac{(\beta-1+2x)(1-x)}{(\beta+1-2x)x} = 0$$

or

$$\left\{ \frac{x}{1-x} \right\}^{(z-2)/z} = \frac{\beta-1+2x}{\beta+1-2x} = C_5. \quad (24.13)$$

This is the equivalent of eqn. (23.7) of the zeroth approximation, and the roots give the equilibrium compositions of the coexisting phases at any temperature. It is possible to eliminate  $\beta$  from (24.13) and thus obtain a more satisfactory expression for the solubility limits. By manipulation of (25.13)

$$\beta^2 - (1-2x)^2 = 4C_5(1-2x)^2/(1-C_5)^2$$

and eliminating  $\beta$  using (24.9)

$$\exp(2\mathcal{E}/kT) = C_5(1-2x)^2/x(1-x)(1-C_5)^2$$

or

$$\exp(\mathcal{E}/kT) = \frac{1-x/(1-x)}{[x/(1-x)]^{1/z} - [x/1-x]^{1-1/z}}. \quad (24.14)$$

The equilibrium conditions are thus given in terms of the ratio  $x/(1-x)$ . The solutions are symmetrical; if one root is  $x_1$ , the other is  $1-x_1$ . This is a necessary consequence of the nearest neighbour model.

At the critical temperature, the two roots coincide at  $x = \frac{1}{2}$ . Substituting  $x = \frac{1}{2}$  into eqn. (24.14) gives an indeterminate equation, so we write  $x/(1-x) = 1 + \delta$ , and by expanding in terms of  $\delta$  and letting  $\delta \rightarrow 0$ ,

$$\exp(\mathcal{E}/kT_c) = z/(z-2). \quad (24.15)$$

For a f.c.c. structure,  $z = 12$  and

$$T_c = \frac{\mathcal{E}}{k} \ln 1.2 = z\mathcal{E}/2.19k,$$

whilst for a b.c.c. structure

$$T_c = \frac{\mathcal{E}}{k} \ln 1.33 = z\mathcal{E}/2.3k.$$

These values may be compared with eqn. (23.10) for the zeroth approximation. If we let  $z \rightarrow \infty$  in eqn. (24.15), we re-obtain the value  $T_c = z\mathcal{E}/2k$ . Guggenheim has emphasized that this is a general result, and any formula of the quasi-chemical method in the first approximation can be converted into the corresponding zeroth order approximation by letting  $z \rightarrow \infty$ .

Equation (24.11) is a much more satisfactory expression for the free energy of a solid solution than is (23.5), especially as it includes a temperature-dependent energy term. In practical applications, however, eqns. (23.7) and (23.10) are usually used in preference to (24.14) and (24.15) because of their greater simplicity. In Fig. 6.9, the predicted solubility

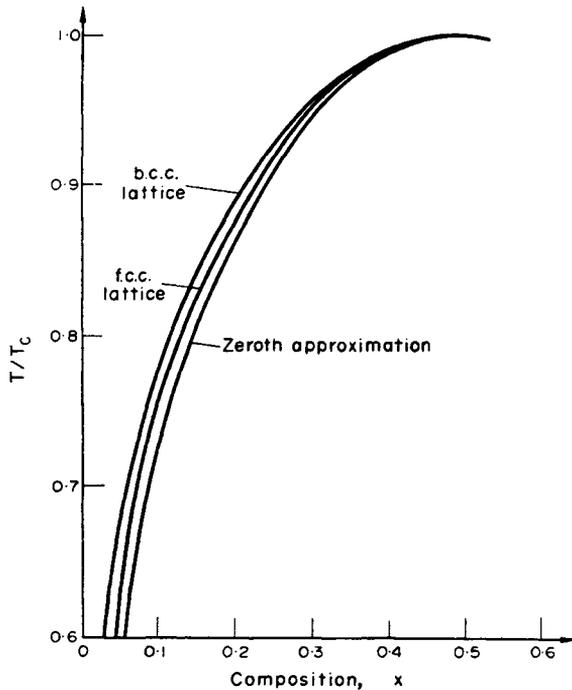


FIG. 6.9. Solubility limits according to the zeroth and first approximations of the quasi-chemical theory.

limits for the first and zeroth approximations are compared by plotting them as a function of  $T/T_c$ . The curve for the first approximation is drawn for both the f.c.c. and the b.c.c. structure. We have assumed identical values of  $T_c$  rather than of  $\mathcal{E}$ , since this most clearly reveals the shape of the curves. In practice,  $T_c$  is measured and  $\mathcal{E}$  is the disposable parameter.

The basic assumption of the method of calculation used in this section is the non-interference of atom pairs. This is a very artificial hypothesis since each atom belongs simultaneously to several pairs. Higher approximations may be obtained by considering the non-interference of interacting groups of larger numbers of sites. All such treatments may be classed as examples of the "cluster variation" method (Kurata, *et al.*, 1953), and the three-dimensional partition function of the Ising model is found approximately by a procedure similar to that outlined above for pairs of sites. A small group of lattice sites is chosen as the basic unit, the energy of this cluster is computed exactly for each way in which the sites may be occupied, and the number of microscopic configurations of the crystal is calculated approximately. The inconsistencies in the calculation of the number of configurations decrease as the size of the basic cluster increases, but the mathematics become correspondingly more complex. Most calculations have been made with triangular clusters of three sites, or with tetrahedral clusters of four sites; in some problems, involving superlattice formation, the choice of the basic cluster is governed by the crystallography in these higher approximations.

We shall not consider further the mathematical development of the Ising model; although the higher approximations lead to better solutions, they are mainly of value in order-disorder problems, where the use of pair interactions sometimes leads to difficulties (see p. 224). The results obtained do not generally differ greatly from those of the zeroth and first approximations; if systems of three sites are considered, for example, the critical temperature is given by

$$T_c = \Xi/k \ln\{(z+1)/(z-3)\}^{1/2},$$

and for a f.c.c. structure, this gives  $T_c = z\Xi/2.21k$ . Table V gives some comparisons of the predictions of the various cluster approximations; it is abstracted from more complete data given by Guggenheim (1952). The value of  $z$  in the cluster approximations is taken to be 12. The values of the excess entropy are negative and very small.

TABLE V. SUMMARY OF PREDICTIONS OF QUASI-CHEMICAL THEORY FOR THERMODYNAMIC FUNCTIONS AT  $T = T_c$

x	Function	Approximation			
		Zeroth	Pairs	Triplets	Quadruplets
0.1	$-\Delta_m g/kT$	0.15	0.13	0.13	0.13
0.3		0.19	0.17	0.17	0.16
0.5		0.19	0.17	0.17	0.17
0.1	$\Delta_m h/kT$	0.18	0.19	0.19	0.19
0.3		0.42	0.42	0.42	0.42
0.5		0.5000	0.50	0.50	0.50
0.1	$-\Delta_c s/k$	0	0.014	0.014	0.015
0.3		0	0.018	0.021	0.023
0.5		0	0.025	0.028	0.033
—	$z\Xi/kT_c$	2	2.19	2.21	2.23

We have already noted that exact solutions of the Ising model have been obtained in two dimensions. For a square lattice, the closed solution is

$$\sinh(z\mathcal{E}/4kT_c) = 1 \quad (24.16)$$

which gives

$$T_c = z\mathcal{E}/3.526k.$$

This formula has been used in Chapter 5, Section 19. Exact solutions in three dimensions have not been found in closed form, but quite early in the development of the theory Kirkwood pointed out that an exact solution for the partition function may be written as a power series in  $\mathcal{E}/kT$ . Unfortunately, this series converges very slowly, and since we are more interested in obtaining a reasonable physical model than in an expression for  $T_c$ , we shall not investigate this further.

In all the above derivations we have assumed that  $\mathcal{E}$  is a constant energy. The non-dependence of  $\mathcal{E}$  on the surrounding configuration of atoms is a fundamental postulate of the quasi-chemical theory, but Guggenheim (1948) has pointed out that there is no need to assume that  $\mathcal{E}$  is independent of temperature. An alternative approach is to regard  $\mathcal{E}$  as a quantity defined by eqn. (24.7). If we do this,  $2\mathcal{E}$  is the free energy increase when an  $A-A$  and a  $B-B$  pair are destroyed to produce two  $A-B$  pairs. This co-operative free energy may now include thermal entropy terms, insofar as these are affected by the interchange, and the restriction to solutions obeying the Kopp-Neumann rule has been removed. Since  $\mathcal{E}$  is now a function of temperature, we may write formally

$$\mathcal{E}' = \mathcal{E} - T(d\mathcal{E}/dT), \quad (24.17)$$

where  $\mathcal{E}'$  corresponds to our old definition of  $\mathcal{E}$  as an interchange energy. The equations for the thermodynamic quantities at any one temperature still contain only one parameter, which may be fixed empirically. On the other hand, the equations for the variation of solubility with temperature are no longer valid, and we cannot deduce  $\mathcal{E}$  or  $\mathcal{E}'$  from experimental data of this kind unless some assumption is made about  $d\mathcal{E}'/dT$ . Guggenheim suggested that the simplest assumption,  $\mathcal{E}'$  independent of temperature, may be adequate.

We now turn to a brief comparison of the predictions of the quasi-chemical theory with experiment. The comparison may be made in various ways; for example, values of  $\mathcal{E}$  can be determined from independently measured quantities and examined for consistency. If the theory is applicable,  $\mathcal{E}$  should also be independent of composition, and the composition dependence of experimentally measured activity coefficients and pair probability functions will show whether or not this is so. Detailed tests of this kind are possible in some assemblies, but are not needed in others, where the signs or the magnitudes of the excess thermodynamic quantities are sufficient to establish that the theory is unsatisfactory.

We note first of all that the formation of a metallic solid solution from its constituents produces fairly small changes in the thermodynamic functions, since the atomic interactions are not greatly changed. Thus, the experimentally measured excess functions are small; the heat of mixing per atom,  $\Delta_m g$ , usually lies within the limits  $\pm kT$ , and the excess entropy of mixing per atom,  $\Delta_m s$ , usually lies within the limits  $\pm k$ . Nevertheless, the magnitude of  $\Delta_m s$  provides one of the most convenient tests of the theory, and especially of the assumption

that the thermal entropy may be neglected. According to the quasi-chemical theory,  $\Delta_e s$  is always negative, and it is readily seen that  $|\Delta_e s/k|$  is unlikely to exceed 0.05 (see Table V). Data collected by Oriani (1959) for a large number of liquid and solid solutions (mainly at  $x = \frac{1}{2}$ ) show that this is rarely true. The experimental values range from  $\Delta_e s/k \approx -1$  for liquid  $Mg_{1/2} Bi_{1/2}$  to  $\Delta_e s/k \approx +0.7$  for solid  $Au_{1/2} Ni_{1/2}$ .

A related test of the assumption about thermal entropy is to consider the equilibrium solubility curves for dilute solid solutions, instead of measured thermodynamic functions. This has been done by Freedman and Nowick (1958) using eqn. (22.27). The partial excess entropies obtained in this way were all positive (with one doubtful exception) and of appreciable magnitude; this is true for assemblies giving  $\Delta_e s_B$  as well as for assemblies in which  $\Delta_e s_B$  (related to  $\Delta_e s_B$  by eqn. (22.26)) is obtained from the equilibrium diagram.

It follows from these results that the assumption of zero thermal entropy of mixing is a serious defect of the theory in most assemblies, and is more important than the error in the configurational term. This means that a theory with  $\bar{E}$  constant is not good enough, and the predictions of the model in respect of temperature variation are inadequate. However, if  $\bar{E}$  is allowed to vary with temperature, the quasi-chemical theory may still be a reasonable representation of the situation in some alloys at constant temperature, and it is, indeed, usually more successful in dealing with the heat of mixing than with the excess entropy.

We end this section by referring briefly to the attempts which have been made to develop better statistical models of solutions. The first of these is the theory of conformal solutions (Longuet-Higgins, 1951), which is a generalization of the pair interaction model, but is free from any structural assumptions. This makes it particularly suitable for dealing with liquid solutions.<sup>†</sup> A conformal solution is defined by certain assumptions about the form of the potential energy of the assembly, expressed as the sum of a number of bimolecular terms. These assumptions, which derive essentially from a law of corresponding states, are that the interaction energies between  $A-A$ ,  $B-B$ , and  $A-B$  pairs depend only on the nature and separation of the two atoms or molecules considered, and have the same functional dependence on the separation for all pairs. Thus the interaction energy of an  $A-A$  pair can be expressed in the form  $\bar{E}_{AA}(r) = f_{AA}\bar{E}_0(g_{AA}r)$ , with equivalent expressions for  $\bar{E}_{BB}(r)$  and  $\bar{E}_{AB}(r)$ . The factors  $f_{AA}$ ,  $f_{AB}$ ,  $g_{AA}$ , etc., are all constants, and  $-\bar{E}_0(r)$  is a universal function giving the potential energy of two atoms or molecules in a reference state as a function of their separation  $r$ . If the reference state is taken to be pure  $A$ , for example,  $f_{AA} = g_{AA} = 1$ .

In developing the theory, it is necessary to assume that  $1-f_{AA}$ ,  $1-g_{AA}$  are both small, so that all the forces have approximately the same magnitude, and it is also assumed that  $g_{AB} = \frac{1}{2}(g_{AA} + g_{BB})$ . The advantage of this approach is that the pair interaction energies vary with composition, since the mean separation is a function of composition, and the difficulties inherent in the usual quasi-chemical assumption of a temperature-independent  $\bar{E}$  are also avoided. The results of the theory are obtained in a series of successive approximations. The averaging procedure in the first approximation leads to a random distribution

<sup>†</sup> The quasi-chemical method is applicable in principle to liquids, and is often used for liquid solutions, but it does carry the implication of a "lattice model" of the liquid.

of the components, and is equivalent to the zeroth approximation of the theory of regular solutions; the higher approximations lead to statistical terms which are very difficult to evaluate (Brown and Longuet-Higgins, 1951).

Prigogine *et al.* (1957) have developed a theory of solutions which we may regard as based on the theory of conformal solutions and on the cell model. They use average interactions between a molecule or atom and its neighbours, and assume composition-dependent potentials of the Lennard-Jones (6-12) type. The interaction potentials are thus written  $-\Xi_{AA}(r) = -\Xi_{AA}^* \varphi(r/r_{AA}^*)$ , etc., where  $\Xi_{AA}^*$ , etc., are characteristic energies, and  $r_{AA}^*$ ,  $r_{AB}^*$ , etc., are characteristic lengths. The function  $\varphi$  is a universal function specifying the type of interaction assumed; for Lennard-Jones forces,  $\varphi(x) = x^{-12} - 2x^{-6}$ . The average potential model assumes random atomic distribution, and the expressions for the thermodynamic functions are obtained in terms of three parameters which are defined by the equations

$$\left. \begin{aligned} \varrho &= (r_{BB}^*/r_{AA}^*) - 1, \\ \delta &= (\Xi_{BB}^*/\Xi_{AA}^*) - 1, \\ \theta &= (\Xi_{AA}^* + \Xi_{BB}^* - 2\Xi_{AB}^*)/\Xi_{AA}^*. \end{aligned} \right\} \quad (24.18)$$

These parameters are essentially "reduced" quantities, and the theory gives expressions for  $\Delta_m h$  and  $\Delta_e s$  which can be simplified considerably when it is justifiable to neglect all products of  $\varrho$ ,  $\delta$ , and  $\theta$  and all powers of higher order than  $\delta$ ,  $\theta$ , and  $\varrho^2$  respectively. Under these conditions, which obviously correspond to those mentioned above in the theory of conformal solutions, we have

$$\left. \begin{aligned} \Delta_m h &\approx 1.435(\theta + 4.5\varrho^2)z\Xi_{AA}^*x(1-x), \\ \Delta_e s &\approx 78.2k\varrho^2x(1-x), \end{aligned} \right\} \quad (24.19)$$

where the factor 1.435 is a result of using Lennard-Jones forces for interactions between atoms which are not nearest neighbours; this factor is unity if only nearest neighbour interactions are considered. The results then reduce to the zeroth approximation of the quasi-chemical theory when  $\varrho = 0$ , and it will be seen that, according to this model, the positive excess entropy which is usually observed is attributable to the size disparity measured by  $\varrho$ .

Shimoji and Niwa (1957) have modified this theory by using generalized interatomic potentials of the Morse type in place of the 6-12 potential. They obtain an expression which is equivalent to Prigogine's equation for  $\Delta_m h$  if the Morse potential is chosen to correspond to Lennard-Jones forces, but they suggest that for liquid metal solutions, a more appropriate choice of potential gives

$$\Delta_m h \approx (\theta + 0.5\varrho^2)z\Xi_{AA}^*x(1-x). \quad (24.20)$$

The predictions of the average potential model have been compared with measured thermodynamic parameters by Oriani (1959). The procedure adopted was to assume  $r_{AB}^* = \frac{1}{2}(r_{AA}^* + r_{BB}^*)$  and to evaluate  $\theta$  from the measured  $\Delta_m g$ . The values of  $\Delta_m h$  and  $\Delta_e s$  were then calculated. The agreement found is rather poor, especially as the theory is only being used to decide how the excess free energy of mixing is distributed between the heat and excess entropy of mixing. The degree of agreement between the experimental and theoretical  $\Delta_m h$  is, moreover, little affected by whether or not the size disparity is taken into account.

The provisional conclusion is thus that the improved statistical theories are little better than the quasi-chemical theory in explaining the properties of solid metallic solutions. The main reason for this situation is that already emphasized on p. 111. Metals and metallic solutions do not obey a law of corresponding states, and the assumption of central force interactions of any type is a poor approximation.

## 25. MISFIT ENERGY IN SOLID SOLUTIONS: SUB-REGULAR SOLUTIONS

We have emphasized the restriction of the quasi-chemical theory to solutions in which the constituents have nearly equal atomic volumes. The failure of the theory in many cases in which large heats of mixing are associated with a considerable difference in the radii of the component atoms is frequently ascribed to "strain energy". This term is seldom precisely defined, but it is usually implied that the energy is a mechanical energy of long-range character (i.e. distributed throughout the crystal), and is somehow separable from any other (chemical) terms in the energy of mixing.

Theories of solid solution which utilize the notion of strain energy attempt to calculate the heat of mixing, or rather the partial excess free energy of mixing, from more fundamental parameters such as elastic constants. All treatments use the approximation of an elastic continuum to represent the solid, and then consider the mechanical strain energy when a solute atom is inserted into this continuum. The model used is one in which a sphere of specified radius is forced into a hole of smaller or larger radius, the surfaces welded together, and the body allowed to relax in its self-stressed state. Each solute atom is thus regarded as a centre of dilatation.

This model seems quite specific, but in fact the calculation has been applied in various ways. The most realistic physical assumption would seem to be to regard the atomic displacements as Hookean outside some limiting radius  $r_0$  from the centre of dilatation, and thus to calculate the strain energy in the region of crystal for which  $r > r_0$ . The total (free) energy of the solute atom, apart from the configurational term, would then be obtained by adding a "core energy" to the elastic energy; this is the approach which is normally used in dislocation theory (Chapter 7). A serious difficulty in the misfitting sphere model is that the elastic energy varies inversely as the cube of  $r_0$ , and so is very sensitive to the choice of this core radius. Different authors have put  $r_0$  equal to the radius of the solute atom itself, the radius of a cluster of twelve nearest neighbours, and the radius of the third nearest neighbour shell.

We should emphasize here that providing the processes envisaged above are isothermal, the calculation gives an estimate of the free energy<sup>†</sup> change associated with the introduction of the solute atom in a particular place, and not the internal energy or the enthalpy. Thus for an isolated solute atom, we are attempting to calculate  $\Delta_e g_B$ , and differentiation with

<sup>†</sup> Strictly, the elastic energy gives the change in the Helmholtz free energy, and the Gibbs free energy is obtained by adding the (negligible) work done against external forces in changing the volume of the assembly. The usual configurational term has, of course, to be added to the elastic energy to give the total free energy change associated with the solute atom.

respect to temperature will give  $\Delta_e s_B$ . This has been mentioned already in connection with Zener's conclusion about the sign of the excess (thermal) entropy of a point defect (p. 129). In some papers it is wrongly implied that the elastic energy is an internal energy.

The usual way of avoiding, or rather concealing, the difficulty of defining  $r_0$  is to apply elastic theory both to the surrounding matrix and to the solute atom and its immediate neighbours. Two energy terms are calculated, one being the work necessary to expand (contract) a hole in the matrix from an initial radius to a new final radius, and the other the work needed to compress (expand) a sphere from its initial radius to the final radius of the hole. The sum of these two contributions is then minimized to give the equilibrium final radius. The initial radii are commonly taken as those of a solvent and solute atom respectively, although a larger cluster may be used with equal justification.

Oriani (1956) has pointed out that this calculation is very unsatisfactory. The minimization procedure of the mechanical model has no real significance in the problem of determining the equilibrium "radius" of the solute atom. It is easy to see that the primary role in determining the distances of the first few coordination shells must be played by electronic interactions (including effects such as charge transfer and screening) with the immediate neighbours of the atom concerned. In a liquid solution, these interactions must entirely determine the "size" of the atom, since there are no long-range elastic stresses. For solid solutions, the same interactions must be present, and will still be the main factor determining the local configuration, but there will be a perturbation resulting from the imposition of lattice periodicity.

In a careful discussion of the significance of misfit energy, Oriani concludes that the term has no operational significance in liquid solutions, and any theoretical concept which might be invoked relates only to hypothetical situations which cannot be obtained physically. (For example, a theoretical definition of misfit energy might be the difference in energy of a real liquid solution and a hypothetical solution in which the components had the same ionic radii; this clearly has no operational significance and the quantity could never be measured.) The idea of size effects and misfit energy should thus be used very cautiously, if at all, for liquid solutions. For solid solutions it is possible to give an operational definition of misfit energy based on the changes produced by the imposition of long-range lattice periodicity. Only that part of the misfit energy outside some (unknown) radius  $r_0$  is susceptible to calculation by elastic theory, and might thus fairly be called strain energy.

These are severe criticisms of the elastic model, and we shall see later that they are to some extent supported by the experimental evidence. The model is capable of giving heats of mixing of the correct magnitude, and entropies of mixing of the correct sign and magnitude, but more detailed tests seem to show that  $\Delta_e g$  cannot really be ascribed to the elastic strain energy in the matrix. Nevertheless, we shall now give the elastic calculation in some detail. The reasons for this are twofold. First, similar calculations are still frequently used in many applications to both solid solutions and lattice defects, and as Eshelby (1956) has remarked, the limitations of this model are perhaps more immediately obvious than the equally serious defects of other approximations used in solid state theory. Second, and more important, we shall require the identical calculation in a later chapter when we consider the energy of a precipitate growing in a matrix. The minimization of the mechanical

strain energy then has much greater validity, and in fact the procedure was first used for this problem by Mott and Nabarro (1940).

In the present application, we consider a solid solution formed from two components of atomic volumes  $v_A^0$  and  $v_B^0$  in the respective pure states. Consider first the effect of introducing a single  $B$  atom into a lattice of pure  $A$ . According to the misfitting sphere model, we treat the  $B$  atom as an elastic sphere inserted into a hole of volume  $v_A^0$  in an isotropic elastic continuum.

Let the atomic radii of the atoms in the pure components be  $r_A^0$  and  $r_B^0 = (1 + \epsilon)r_A^0$ , where  $\epsilon$  may be positive or negative. We suppose first that the  $B$  atom is introduced into an infinite  $A$  crystal, where its volume becomes  $v_B$  and its effective radius is  $r_B = (1 + C_6\epsilon)r_A^0$ . Use an origin at the centre of the  $B$  atom. From the symmetry of the problem it follows that the displacements  $\mathbf{w}$  are all radial, and  $|\mathbf{w}|$  is a function only of the radius vector  $r$ . In polar coordinates, the strain components are

$$e_{rr} = \partial w / \partial r, \quad e_{\theta\theta} = e_{\phi\phi} = w/r,$$

and hence

$$\Delta = (\partial w / \partial r) + 2w/r, \quad \boldsymbol{\omega} = \text{curl } \mathbf{w} = 0. \quad (25.1)$$

The equations of equilibrium in the form (11.30) thus reduce to

$$(\lambda + 2\mu) \{(\partial^2 w / \partial r^2) + (2/r)(\partial w / \partial r) - 2w/r^2\} = 0,$$

and the general solution is

$$w = A_1 r + A_2 / r^2.$$

The displacements in any spherically symmetrical elastic problem must be of this type. In our infinite medium, we have to satisfy the boundary conditions

$$w(r_A^0) = C_6 \epsilon r_A^0, \quad w(\infty) = w(0) = 0, \quad (25.2)$$

if the initial state is taken to be the introduction of the compressed  $B$  atom into the hole of radius  $r_A^0$ . The displacements are thus

$$\left. \begin{aligned} w'_B &= C_6 \epsilon r & (r < r_A^0), \\ w_A &= C_6 \epsilon (r_A^0)^3 / r^2 & (r > r_A^0). \end{aligned} \right\} \quad (25.3)$$

The vector  $\mathbf{w}_A = C_6 \epsilon \{(r_A^0)^3 / r^3\} \mathbf{r}$  gives the displacements in the  $A$  crystal relative to the unstressed state. The chosen initial state for the  $B$  atom, however, is when it has been compressed or extended from radius  $(1 + \epsilon)r_A^0$  to  $r_A^0$ . The displacements  $w'_B$  refer to this state, and relative to the unstressed  $B$  sphere (before its introduction into the hole), we have the displacements

$$w_B = \epsilon(C_6 - 1)r,$$

so that within this sphere  $e_{rr} = e_{\theta\theta} = \Delta_B / 3 = (C_6 - 1)\epsilon$ . The state of stress is a uniform hydrostatic pressure (or tension) given by

$$p = 3K_B(C_6 - 1)\epsilon, \quad (25.4)$$

where  $K_B$  is the effective bulk modulus of the solute atom  $B$ . It is usual in this model to take  $K_B$  as the bulk modulus of the component  $B$ . This is admittedly a poor approximation because of the changes in the electronic state of a strongly deformed atom.

The introduction of the  $B$  atom leaves the whole assembly in a state of self-stress, and the stress components in the surrounding matrix are  $X_{rr}$  and  $X_{\theta\theta} = X_{\varphi\varphi}$ . From (25.1) and (25.3) we see that  $\Delta_A = 0$ , so that only shear strains are present, and the stresses may be written entirely in terms of the shear modulus  $\mu_A$  of  $A$ . Using (11.26) the stress components are

$$\left. \begin{aligned} X_{rr} &= 2\mu_A(\partial w/\partial r) = -4\mu_A C_6 \varepsilon(r_A^0)^3/r^3, \\ X_{\theta\theta} &= 2\mu_A w/r = 2\mu_A C_6 \varepsilon(r_A^0)^3/r^3. \end{aligned} \right\} \quad (25.5)$$

Since the matrix and inserted sphere are in equilibrium,  $X_{rr}$  must equal  $p$  (eqn. (25.4)) when  $r = r_B \approx r_A^0$ . This gives an expression for  $C_6$ :

$$C_6 = 3K_B/(3K_B + 4\mu_A). \quad (25.6)$$

If the  $B$  atom is incompressible ( $K_B = \infty$ ),  $C_6 = 1$  and  $v_B = v_B^0$ .

Let  $S$  be any closed surface in the assembly which totally encloses the  $B$  atom. An element  $dO$  of this surface, having unit vector normal  $\mathbf{n}$ , moves when the  $B$  atom is introduced so as to sweep out a volume  $\mathbf{w}_A \cdot \mathbf{n} dO$ . The volume within  $S$  thus increases by

$$\left. \begin{aligned} \Delta v^\infty &= C_6 \varepsilon(r_A^0)^3 \int_S (\mathbf{r} \cdot \mathbf{n}/r^3) dO \\ &= 4\pi C_6 \varepsilon(r_A^0)^3 \end{aligned} \right\} \quad (25.7)$$

since the integral is just the total solid angle subtended by  $S$  at the origin. We write the change of volume  $\Delta v^\infty$  to emphasize that this equation is only valid in an infinite crystal. Note that  $\Delta v^\infty$  is equal to the increase in volume of the hole, and is independent of  $S$ , as is necessary since there is no dilatation outside the  $B$  atom. If we write the difference in atomic volumes of the two pure components as  $\Delta v_{AB} = v_B^0 - v_A^0 = 4\pi \varepsilon(r_A^0)^3$ , we also have

$$\Delta v^\infty = C_6 \Delta v_{AB}. \quad (25.8)$$

Now consider what happens in a finite bounded crystal of  $A$ . The boundary conditions (25.2) are no longer applicable, and must be modified so that the external surface is free of traction. This may be accomplished by superimposing on the displacements (25.3) a second set of "image" displacements, caused by surface tractions which will just annul those given by (25.5). The work of Eshelby (1954, 1956) has shown that these image effects are often of surprising importance in the continuum theory of lattice defects.

Suppose the external surface is  $S$ . Then the image displacements are due to tractions  $-X_{ij}n_j$  distributed over  $S$ . If  $S$  is a sphere of radius  $R$ , the surface traction to be applied is a uniform hydrostatic tension (or pressure) of  $4\mu_A C_6 \varepsilon(r_A^0)^3/R^3$ . The corresponding image displacements represent a uniform dilatation of the large sphere provided we neglect the small perturbation caused by the different elastic constants of the enclosed  $B$  atom. This dilatation is

$$\Delta_A = 4\mu_A C_6 \varepsilon(r_A^0)^3/(K_A R^3), \quad (25.9)$$

and the increase in volume of the whole assembly is

$$\left. \begin{aligned} \Delta v^i &= 16\pi\mu_A C_6 \varepsilon (r_A^0)^3 / 3K_A \\ &= (4\mu_A / 3K_A) \Delta v^\infty. \end{aligned} \right\} \quad (25.10)$$

Since the dilatation varies as  $1/R^3$ ,  $\Delta v^i$  is independent of  $R$ . Thus for any sphere, the total change in volume is

$$\left. \begin{aligned} \Delta v &= \Delta v^\infty + \Delta v^i = 4\pi C_6 \varepsilon (r_A^0)^3 / C_6' \\ &= (C_6 / C_6') \Delta v_{AB}, \end{aligned} \right\} \quad (25.11)$$

where  $C_6' = 3K_A / (3K_A + 4\mu_A)$ .

If the bulk moduli of the matrix and the  $B$  atom are equal,  $C_6 = C_6'$  and  $\Delta v = \Delta v_{AB}$ . This illustrates the importance of including the image effects, since this result cannot be otherwise obtained.

A solution for the image displacements is possible only when the surface  $S$  has certain simple forms. However, it is readily proved (Eshelby, 1956) that the result for the change in volume is valid for any external surface. In a sphere with a uniform distribution of solute atoms,  $c_B$  per unit volume, it follows from symmetry that the shape of the assembly is unaltered, apart from small ripples in the surface, of magnitude determined by the mean separation of  $B$  atoms. The fractional change in the volume of the assembly will be

$$\Delta V / V = c_B (C_6 / C_6') \Delta v_{AB}. \quad (25.12)$$

Eshelby has shown that this result remains true for a body of arbitrary shape which contains a uniform distribution of defects.

Returning to the single  $B$  atom in otherwise pure  $A$ , we now calculate the strain energy. Within the compressed or extended  $B$  sphere, the strain energy density,  $\frac{1}{2}p\Delta$ , is constant, and the total strain energy is thus

$$W_B = 9v_B K_B (C_6 - 1)^2 \varepsilon^2 / 2. \quad (25.13)$$

If the atom is in an infinite matrix, the strain energy density at any point is

$$\frac{1}{2} X_{rr} e_{rr} + X_{\theta\theta} e_{\theta\theta} = 6\mu_A C_6^2 \varepsilon^2 (r_A^0)^6 / r^6,$$

and the total strain energy by integration is

$$W_A = 6\mu_A C_6^2 \varepsilon^2 v_B \quad (25.14)$$

to the first order in which all these equations are valid. The same result for  $W_A$  is obtained by considering the strain energy density in a finite medium and integrating from  $r_B$  to  $R$ .

The strain energy of the whole assembly is thus

$$W_s = W_A + W_B = 6\mu_A C_6 \varepsilon^2 v_B \quad (25.15)$$

using the result of (25.6). Since  $\varepsilon = \Delta v_{AB} / 3v_B$ , we may also write the strain energy in the form

$$W_s = 2\mu_A C_6 (\Delta v_{AB})^2 / 3v_B. \quad (25.16)$$

Note that the volume  $v_B$  in eqns. (25.13), (25.14), and (25.16) might equally well be written  $v_A^0$  or  $v_B^0$  to the order of approximation of linear elasticity theory, and it would perhaps be better to write these equations with some mean volume  $v$ .

We can now consider the properties of a solid solution containing  $N_A$   $A$  atoms and  $N_B$   $B$  atoms. If  $N_B$  is small, the volume of the assembly will be  $V + N_B \Delta v$ , where  $\Delta v$  is given by eqn. (25.11). The mean volume per atom is thus

$$\overline{v_{AB}} = v_A^0 + x \Delta v,$$

where  $x$  is the atomic concentration of  $B$  atoms. Substituting  $\Delta v = (C_6/C_6')(v_B^0 - v_A^0)$ , we find

$$\overline{v_{AB}} = (1-x)v_A^0 + xv_B^0 + x(C_6/C_6' - 1)(v_B^0 - v_A^0). \quad (25.17)$$

When  $K_A = K_B$ , this leads to an additive law for atomic volumes, or on the linear approximation to an additive law for atomic radii

$$\overline{r_{AB}} = (1-x)r_A^0 + xr_B^0 = (1 + \varepsilon x)r_A^0, \quad (25.18)$$

so that  $\varepsilon$  is the fractional rate of change of lattice constant  $\partial r/\partial x$ . Equation (25.18) is usually known as Vegard's law, and according to the elastic model the radii are larger or smaller than their Vegard's law values according to whether the last term in (25.17) is positive or negative. The sign of the deviation is thus dependent on the sign of

$$(K_B - K_A)(r_B^0 - r_A^0).$$

Friedel (1955) has shown that this rule accounts for observed qualitative deviations from the law (using X-ray measurements), although quantitative agreement, as expected, is not very good.

In deriving eqn. (25.17) we have assumed that  $\Delta v$  is a constant independent of  $x$ , and we have also neglected the variation of elastic constants with  $x$ . It might strictly be more logical to assume that the  $B$  atom is forced into a hole of size  $v_A^0(1 + \Delta_A)$ , where  $\Delta_A$  is the dilatation produced in the matrix by all the preceding  $B$  atoms. However, the misfitting sphere model is probably too uncertain in its details for this kind of modification. Eshelby has pointed out that there is no very convincing reason to take the volumes of the hole and misfitting sphere as  $v_A^0$  and  $v_B^0$  respectively; they might, for example, have been chosen so that their radii were equal to nearest neighbour distances in the pure components. Experimental results which show that Vegard's law is approximately valid in many alloys are the best justification for the assumption we have made.

We are now able to derive the strain energy of the solid solution. When the composition is  $x$ , the addition of a further  $B$  atom changes the energy first by the self-energy  $W_s$  of this atom, and, secondly, by the interaction energy of the new solute atom with all the preceding  $B$  atoms. This interaction energy is due to the presence of a hydrostatic tension or compression in the matrix (the image stress), and as this helps the required expansion or contraction of the hole, the total extra energy is less than  $W_s$ . Although the interaction energy of any two  $B$  atoms is negligible in comparison with  $W_s$ , the fact that each  $B$  atom interacts with

all the other  $B$  atoms means that the total interaction energy is comparable with the total self-energy.

When the concentration is  $c_B$ , the dilatation is  $c_B \Delta v^i = (x/v_A^0) \Delta v^i$ . The interaction energy when another  $B$  atom is added is thus  $K_A(x/v_A^0) \Delta v^i \Delta v = 2W_s(C_6/C_6')x \simeq 2W_s(K_B/K_A)x$ . The net change in energy on adding the  $B$  atom is thus

$$\delta W = W_s \{1 - 2(K_B/K_A)x\},$$

and this corresponds to a change of composition  $\delta x = 1/N$ . The total strain energy is thus

$$W = NW_s x \{1 - (K_B/K_A)x\} \quad (x \ll 1). \quad (25.19)$$

If chemical interaction terms are small, this strain energy is the atomic part of the free energy of solution, as pointed out on p. 198, and the total free energy of solution is obtained by adding the configurational term for random mixing to give

$$\Delta_m g = W_s x \{1 - (K_B/K_A)x\} + kT \{x \ln x + (1-x) \ln(1-x)\} \quad (x \ll 1). \quad (25.20)$$

Except for the factor  $(K_B/K_A)$ , this equation is identical in form with eqn. (23.5) for a regular solution. However, we have already emphasized that  $W$  is a free energy, so that the temperature variation of the elastic constants gives rise to an appreciable temperature variation of  $W$ , and hence to an entropy

$$\Delta_e s = -(\partial W / \partial T). \quad (25.21)$$

From eqn. (25.16) we see that the temperature variation of  $W$  is governed by the term  $\partial \mu / \partial T$ , and since this is always negative, the elastic model always gives a positive value to  $\Delta_e s$  (Zener, 1951). Friedel (1955) has compared the values of  $\Delta_e s$  and  $\Delta_m h$  obtained from eqns. (25.21) and (25.19) with experimental data on gold-nickel alloys, and obtained good agreement. In a survey of limiting solubility data for a number of binary alloys, using the method of analysis described on p. 200, Freedman and Nowick (1958) found that  $\Delta_e s$  is always positive, in accordance with the predictions of the model. The model actually permits calculation of  $\Delta_e s$ , but, as pointed out earlier, it is too sensitive to a choice of integration limits for this calculation to have much significance. The values obtained are certainly of the correct magnitude, but it is also reasonable to expect some proportionality between the observed  $\Delta_e s$  and that calculated from the size factor and temperature variation of the shear modulus. Freedman and Nowick showed that such a proportionality does not exist, and concluded that a quantitative interpretation of  $\Delta_e s$  cannot be given by the elastic model. Such an interpretation almost certainly requires detailed consideration of nearest neighbour interactions, as in Huntington's calculations for point defects (see p. 130). Oriani (1959) reaches the same conclusion, which he bases partly on the observed experimental correlation between heats of solution in the liquid and solid phases of alloys with a size disparity. The elastic model is not applicable to a liquid solution, so there should be no correlation if it is valid.

Within the limitations of the model, the equations derived above are strictly valid only for very dilute solutions. Some attempts have been made to treat more concentrated solutions by supposing that each atomic site is occupied by a positive or negative strain centre

acting in a mean lattice. We shall not describe these theories because of their complexity; some attention has been paid to the X-ray diffraction effects expected from a model of this type. The simplest treatment of more concentrated solutions comes from noting that eqn. (25.20) will be nearly identical for both *A*-rich and *B*-rich solutions if the components have nearly identical elastic properties, and as a first approximation may be applied to intermediate compositions, using mean values for the parameters. Comparison with the previous quasi-chemical treatment then shows that the critical temperature below which phase segregation occurs is given by

$$T_c = 3\bar{C}_8 \varepsilon^2 \bar{v} \bar{\mu} / k,$$

and if this is to be below the melting point of the alloy,

$$\varepsilon^2 < (kT^s / 3\bar{C}_8 \bar{v} \bar{\mu}).$$

Reasonable values of the constants give  $\varepsilon \sim 0.15$ , in agreement with the empirical Hume-Rothery rule that wide solid solution is possible only when the size factor  $\varepsilon$  is less than about 15%. In view of the reservations above, it is doubtful if this is anything other than a coincidence.

An early attempt at a semi-empirical equation for the strain energy was made by Lawson (1947). He used the expression for the self-energy of a *B* atom in an *A* lattice, but replaced  $\Delta v_{AB}$  by  $v_B^0 - v_{AB}$ . This means that the hole into which the *B* atom is forced is equal in volume to the mean volume per atom in the assembly. Lawson neglected interaction energies, and his equation for the free energy may be expressed in the form

$$\Delta_m g = A_1 x(1-x)^2 + A_2 x^2(1-x) + kT \{x \ln x + (1-x) \ln(1-x)\}. \quad (25.22)$$

The assumptions made in deriving the equation are too sweeping for the expressions for  $A_1$  and  $A_2$  to have any real value, but the equation itself may be useful as a representation of the properties of some real solutions. Its real advantage as an empirical equation is that it effectively replaces the constant  $\bar{E}$  of the quasi-chemical theory by a composition dependent exchange energy, and the assumed independence of the quantities  $\bar{E}_{AA}$ ,  $\bar{E}_{AB}$ , etc. of composition is a major weakness of the quasi-chemical theory.

The formal properties of eqn. (25.22) have been investigated by Hardy (1953), who used it to define a "sub-regular solution". For the present, we assume that  $A_1$  and  $A_2$  are independent of temperature and composition, and the chemical potentials are then

$$g_A - g_A^0 = x^2(2A_1 - A_2) + 2x^3(A_2 - A_1) + kT \ln(1-x). \quad (25.23)$$

The compositions of the solubility curve at any temperature are found by differentiating (25.22) and applying the condition of eqn. (22.5). This gives

$$kT \ln \left\{ \frac{x_2(1-x_1)}{x_1(1-x_2)} \right\} + (x_2 - x_1) \{2(A_2 - 2A_1) + 3(A_1 - A_2)(x_1 + x_2)\} = 0, \quad (25.24)$$

where, as before,  $x_1$ ,  $x_2$  are the limits of solubility of *B* in *A* and *A* in *B* respectively. The equation of the spinodal is similarly found to be

$$2(A_2 - 2A_1) + 6x(A_1 - A_2) + kT/x(1-x) = 0, \quad (25.25)$$

and at the critical point

$$\partial^3 \Delta_m g / \partial x^3 = 6(A_1 - A_2) + (2x - 1) kT / x^2(1 - x)^2 = 0. \quad (25.26)$$

From eqns. (25.25) and (25.26) the critical temperature  $T_c$  is related to the critical composition at which the solubility gap closes,  $x_c$ , by the equations

$$\left. \begin{aligned} A_1 &= kT_c(-9x_c^2 + 8x_c - 1)/6x_c^2(1 - x_c)^2. \\ A_2 &= kT_c(-9x_c^2 + 10x_c - 2)/6x_c^2(1 - x_c)^2. \end{aligned} \right\} \quad (25.27)$$

From eqns. (25.27) we see that if  $A_1 = A_2$ ,  $x_c = \frac{1}{2}$  and  $T_c = A_1/2k$ . The model is then formally equivalent to the quasi-chemical model. As  $A_1 - A_2$  increases, the composition of the maximum in the solubility gap moves further away from the equi-atomic composition, and the asymmetry of the solubility gap increases.

Instead of using eqn. (22.5), we could have obtained (25.24) from the equivalent thermodynamic conditions

$$g_A(x_1) = g_A(x_2), \quad g_B(x_1) = g_B(x_2).$$

If we multiply the first of these equations by  $2 - (x_1 + x_2)$  and the second by  $(x_1 + x_2)$  and then add, we obtain

$$(2 - x_1 - x_2) kT \ln\{(1 - x_1)/(1 - x_2)\} + (x_1 + x_2) kT \ln(x_1/x_2) = -(A_1 - A_2)(x_2 - x_1)^3. \quad (25.28)$$

In a regular solution, both sides of this equation are zero. Hardy used (25.28) to test the sub-regular solution model for a number of alloy assemblies. Provided  $(A_1 - A_2)$  is independent of temperature, a straight line may be obtained by plotting the left-hand side of the equation (obtained from the experimental equilibrium diagram) against  $(x_2 - x_1)^3$ . In this way, the alloys Ag-Cu, Ag-Pt, Al-Zn, and Au-Pt were shown to behave approximately as sub-regular solutions. The simple model could not be fitted to other alloys, e.g. Au-Fe, Au-Co, and Au-Ni, where  $A_1 - A_2$  is not constant and  $A_1$  and  $A_2$  are apparently functions of composition. In the alloys which may be represented as sub-regular solutions,  $A_1$  and  $A_2$  may increase with temperature, although the difference  $A_1 - A_2$  is nearly independent of temperature.

The theory of sub-regular solutions does not rest on any firm basis, and has very limited utility. It supplies an empirical equation for the solubility curve in terms of a single parameter  $(A_1 - A_2)$  (eqn. (25.28)), and is thus useful in representing the properties of alloys with equilibrium diagrams which are not symmetrical about the equi-atomic composition.

## 26. ORDERED STRUCTURES IN ALLOYS

In the previous sections we have concentrated mainly on deriving the equilibrium conditions for an assembly with a positive heat of mixing, i.e. an assembly in which  $\bar{E}$  is positive according to the nearest neighbour model. As already noted, a negative  $\Delta_m h$  (negative  $\bar{E}$ ) implies that atoms of opposite kinds attract each other. At low temperatures, each  $A$  atom will thus surround itself with as many  $B$  atoms as possible; such a structure is called a superlattice. A completely ordered structure of this kind resembles a chemical compound, and

is possible only when the atomic fractions of the different components are small integral numbers.

The term superlattice occasionally causes some confusion. As described on p. 171, we do not need to distinguish between the different kinds of atom in a substantially disordered solid solution, since they occupy the available sites approximately at random. When the structure is ordered, the atoms segregate in such a way that the atoms of one kind occupy one or more sets of sites, and atoms of another kind occupy different sets of sites. A distinction must then be made between vectors specifying the positions of *A* atoms and those specifying the positions of *B* atoms, and the size of the primitive unit cell of the Bravais lattice has to be increased. This is the origin of the term superlattice. The conventional unit cell may remain the same size, as happens, for example, in the simplest type of superlattice, which is formed from a disordered b.c.c. structure. If there are equal numbers of *A* and *B* atoms, they may arrange themselves so that all the *A* atoms are at cube corners and all the *B* atoms at cube centres. The superlattice unit cell is the same as the cubic unit cell of the b.c.c. structure, but the structure is now simple cubic because of the non-equivalence of the corner and centre atoms. The cubic unit cell is the smallest possible for the superlattice structure, but the disordered structure has a smaller rhombohedral unit cell. In other superlattices, more complex changes take place, and the formation of the superlattice may be accompanied by a lowering of lattice symmetry.

It would be inappropriate in this book to discuss the structural features of all the known superlattices, but it is convenient to summarize the main types. In binary alloys, most superlattices have one of five structures, two being derived from each of the common cubic structures and one from the h.c.p. structure. In addition, there are a number of superlattice structures of which only one example is known, and some structures which have large unit cells, corresponding to a modification of one of the basic types.

The superlattices derived from the f.c.c. structure (*A1*) are known as  $L1_0$  and  $L1_2$  in the Strukturbericht notation. The unit cell of each contains the four atoms found in the cubic unit cell of the disordered *A1* structure, but  $L1_0$  has tetragonal symmetry. In this structure, which occurs at equi-atomic positions, the *A* atoms are at points [000] and  $[\frac{1}{2} \frac{1}{2} 0]$  of the unit cell, and the *B* atoms at points  $[\frac{1}{2} 0 \frac{1}{2}]$  and  $[0 \frac{1}{2} \frac{1}{2}]$ . The structure consists of alternate layers of *A* and *B* atoms parallel to the (001) planes. The attraction between *A* and *B* atoms results in slightly smaller interatomic distances between nearest neighbours in adjacent layers, so the structure is tetragonal with *c/a* slightly smaller than unity. Each atom has four nearest neighbours of its own type in the same (001) layer, and eight nearest neighbours of opposite type in the two adjacent (001) layers. This contrasts with the completely disordered structure, where each atom on the average has six like and six unlike nearest neighbours.

The  $L1_2$  structure corresponds to the ideal composition  $A_3B$ . The *B* atoms are in the [000] positions, and the *A* atoms in the remaining positions of the conventional unit cell of the f.c.c. structure. In the superlattice, the *B* atoms each have twelve unlike nearest neighbours, compared with an average of three like and nine unlike nearest neighbours in a random f.c.c. solid solution. There is also a single known example of the  $L1_1$  structure, in which alternate (111) planes are composed entirely of *A* and *B* atoms respectively.

The simplest superlattice ( $B2$  or  $L2_0$ ) derived from the b.c.c. ( $A2$ ) structure is found in equi-atomic alloys, and was described above. The symmetry is simple cubic, and the unit cell contains the two atoms of the conventional b.c.c. unit cell. The  $A$  atoms are at the corners  $[000]$  and the  $B$  atoms at the body centres  $[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$ , or vice versa; this is sometimes called the caesium chloride structure. Each atom has eight unlike nearest neighbours in the disordered state compared with an average of four like and four unlike neighbours in the completely disordered state. Another superlattice derived from the b.c.c. structure is more complex; it is known as the  $D0_3$  type. This structure has a cubic unit cell formed from eight conventional b.c.c. unit cells, and thus containing sixteen atoms; the ideal composition is  $A_3B$ . The corner positions of the small cubic cells are occupied by equal numbers of the atoms of each kind, each set being arranged on tetrahedral groups of sites. The body centred positions of the small unit cells are occupied entirely by  $A$  atoms. Each  $B$  atom has eight unlike nearest neighbours in the superlattice, compared with an average of two like and six unlike nearest neighbours in the substantially disordered solid solution.

From the h.c.p. structure ( $A3$ ), a superlattice is made by stacking together four unit cells of two atoms to give a larger unit cell of sides  $2a, 2a, c$  containing eight atoms. The composition is again  $A_3B$ , and each close packed layer contains three times as many  $A$  atoms as  $B$  atoms. The  $B$  atoms form a hexagonal network of side  $2a$ , and the  $A$  atoms occupy the remaining sites of the hexagonal network of side  $a$ . Each  $B$  atom is surrounded by twelve  $A$  atoms in the superlattice compared with an average of three like and nine unlike atoms in the random solution.

The relations we have just described are summarized in Table VI, which also lists some of the known binary assemblies having superlattices of these kinds. In some cases, the

TABLE VI. COMMON SUPERLATTICE STRUCTURES

Disordered structure	Superlattice type	Composition	Atom positions	Examples
f.c.c.	$L1_0$ (tetragonal)	$AB$	$2A$ in $(000; \frac{1}{2}, \frac{1}{2}, 0)$ $2B$ in $(\frac{1}{2}, 0, \frac{1}{2}; 0, \frac{1}{2}, \frac{1}{2})$	AuCu, CoPt, MgIn MnNi, NiPt, FePd, FePt
f.c.c.	$L1_2$ (cubic)	$A_3B$	$3A$ in $(0\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, 0, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 0)$ $1B$ in $(000)$	$Cu_3Au, Au_3Cu,$ $Pt_3Co, Fe_3Pt, Pt_3Fe,$ $Cu_3Pt, Ni_3Mn, etc.$
b.c.c.	$B2 (L2_0)$ (cubic)	$AB$	$1A$ in $(000)$ $1B$ in $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$CuZn, CuPd, AgCd,$ $AgZn, CoFe$
b.c.c.	$D0_3$ (cubic, face-centred)	$A_3B$	$(000; 0\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, 0, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 0)$ $+ 4B$ in $(000)$ $4A$ in $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ $8A$ in $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}; \frac{3}{4}, \frac{3}{4}, \frac{3}{4})$	$Fe_3Al, Fe_3Si,$ $Mg_3Li, Cu_3Al$
h.c.p.	$D0_{19}$ (hexagonal)	$A_3B$	$6A$ in $(\frac{1}{2}, 0, 0; 0, \frac{1}{2}, 0; \frac{1}{2}, \frac{1}{2}, 0);$ $\frac{1}{6}, \frac{1}{3}, \frac{1}{2}; \frac{1}{6}, \frac{5}{6}, \frac{1}{2}; \frac{2}{3}, \frac{5}{6}, \frac{1}{2})$ $2B$ in $(000, \frac{2}{3}, \frac{1}{3}, \frac{1}{2})$	$Ag_3In, Mn_3Ge,$ $Mg_3Cd, Cd_3Mg,$ $Ni_3Sn$

nearest neighbour model with negative  $\bar{\mathcal{E}}$  predicts that for a given crystal structure and composition the lowest energy state will correspond to an observed superlattice type. However, this is not always valid, and the model may be ambiguous; for example, the energy of a b.c.c. solution is reduced by the formation of a phase mixture of the  $L2_0$  superlattice and pure  $A$ , so that all  $B$  atoms have only  $A$  atoms as nearest neighbours. For an  $AB_3$  alloy it follows that the energy of this phase mixture is identical with that of a  $DO_3$  superlattice, so that the nearest neighbour model is not able to predict which is the equilibrium state. It then becomes necessary to consider higher neighbour interactions; this procedure is not so artificial as first appears since, as previously noted, the interchange energies can be given operational definitions (Clapp and Moss, 1968).

Several early calculations of the lowest energy states with second and sometimes third nearest neighbour interactions taken into consideration were restricted either to stoichiometric compositions or to one-phase states. A systematic investigation for cubic alloys with first and second neighbour interactions was first made by Richards and Cahn (1971) and was supplemented by Allen and Cahn (1972). The results are expressed in terms of the ratio  $\xi = \bar{\mathcal{E}}_2/\bar{\mathcal{E}}_1$  (see eqn. (23.1a)) which is unrestricted in sign and magnitude; Richards and Cahn considered only  $\bar{\mathcal{E}}_1 < 0$ , but Allen and Cahn also treated  $\bar{\mathcal{E}}_1 > 0$ . Because of symmetry, it is only necessary to consider explicitly the composition range  $0 \leq x \leq \frac{1}{2}$ .

The internal energy of any configuration may now be expressed in an obvious extension of eqn. (24.1) as

$$\begin{aligned} \Delta_m U &= z_1 N_{AB,1} \bar{\mathcal{E}}_1 + z_2 N_{AB,2} \bar{\mathcal{E}}_2 \Big\} \\ &= \frac{1}{2} N \bar{\mathcal{E}}_1 [z_{AB,1} + z_{AB,2} \xi], \Big\} \end{aligned} \quad (26.1)$$

where  $z_{AB,i}$  now denotes the average number per atom of  $i$ th neighbours of opposite type. Thus for  $\bar{\mathcal{E}}_1$  negative the equilibrium state at 0 K (ground state) is found by maximizing the quantity in square brackets, whilst for  $\bar{\mathcal{E}}_1$  positive the energy will be minimized by minimizing this quantity. There are limits on  $z_{AB,1}$  and  $z_{AB,2}$ , given for  $0 \leq x \leq \frac{1}{2}$  by

$$0 \leq z_{AB,i} \leq 2z_i x. \quad (26.2)$$

The lower limit corresponds to all  $B$  atoms having only  $B$  atoms for  $i$ th neighbours; one way of attaining this is to postulate a two-phase mixture of the two components in which  $z_{AB,i} = 0$  for all  $i$ . At the upper limit, each  $B$  atom has only  $A$  atoms for  $i$ th neighbours, assuming  $B$  to be the minority component.

Clearly, if  $z_{AB,1}$  and  $z_{AB,2}$  can have the upper or lower limiting values (depending on the signs of  $\bar{\mathcal{E}}_1$  and  $\bar{\mathcal{E}}_2$ ) which will minimize  $\Delta_m U$ , a ground state has been found. Thus if both  $\bar{\mathcal{E}}_1$  and  $\bar{\mathcal{E}}_2$  are positive, minimum energy is given by the phase mixture ( $\Delta_m U = 0$ ) as already concluded. However, the crystal structure does not always allow the upper limits of (26.2) to be attained; for example  $z_1 = 12$  for a f.c.c. structure, but the maximum value of  $z_{AB,1}$  for an equi-atomic superlattice is 8. This arises because in this structure two nearest neighbours of a given atom may also be nearest neighbours of each other, and the upper limit of (26.2) is in fact only possible for  $0 \leq x \leq 0.25$ . In addition to this difficulty,  $z_{AB,1}$  and  $z_{AB,2}$  may not be independent of each other, so that it may be impossible to

find a configuration which produces the appropriate limits simultaneously, even though both may be individually obtained.

In cases where  $B-B$  bonds cannot be avoided for negative  $\bar{\mathcal{E}}_i$ , or  $A-B$  bonds cannot be avoided for positive  $\bar{\mathcal{E}}_i$ , the equilibrium structure(s) can only be found by calculating the energies of various possible configurations. Richards and Cahn used an empirical procedure of examining superlattices with translational vectors that coincide as much as possible with energetically favourable distances for like atoms. The results of this procedure seem intuitively to be correct, but no proof could be given that all possible low energy configurations had been considered. Allen and Cahn used a more rigorous technique, the cluster method, that includes consideration of all possible structures but also sometimes includes cluster combinations which represent impossible "structures".

In the cluster method, a motif of  $M$  adjacent lattice points (atoms) is defined by a specific three-dimensional circuit and the  $2^M$  clusters which correspond to the various occupancies of these sites by  $A$  and  $B$  atoms are enumerated. An energy can now be assigned to each cluster and the total energy of any arrangement of  $A$  and  $B$  atoms on a lattice can then be expressed as a linear sum of each such energy multiplied by the number of clusters of that type. The problem of minimizing this energy function subject to the constraints imposed by the overall composition can be solved by a mathematical technique known as linear programming and yields as a result the fractional numbers of clusters of each type in the whole arrangement. This gives an absolute minimum to the energy since all possible structures (single and multiphase) are represented. However, whilst the fractions of each cluster in any arrangement can always be specified, it is not always possible to construct an arrangement corresponding to specified fractions. Thus the numbers given by the technique may represent an "imaginary" structure, and the method then gives only a lower limit to the energy unless a true minimum energy structure can be found by selection of a different motif.

In the case of the b.c.c. structure, the results obtained by the cluster method for negative  $\bar{\mathcal{E}}_1$  coincide with those found by Richards and Cahn, and thus constitute a proof that their ground state diagram is correct. The equilibrium phases at 0 K are shown schematically in Fig. 6.10 in which the phase fields are plotted as functions of  $\xi = \bar{\mathcal{E}}_2/\bar{\mathcal{E}}_1$  and of atomic per cent of  $B$ . The only structure not listed in Table VI is the  $B32$  (NaTl) type. For  $\xi$  negative a two-phase state is stable over the whole composition range, but for  $\xi$  positive the situation is more complex. For  $0 \leq \xi \leq \frac{2}{3}$  and 25–50 atomic %  $B$ , for example the same energy is obtained either by a phase mixture of the two stoichiometric phases,  $DO_3$  and  $B2$ , or by a gradual change in the occupancy of some of the sites. In the latter case, there is a continuous single phase region which includes the stoichiometric compositions. This degeneracy in the ground state is presumably a property of the particular model, and would disappear in a better approximation. Richards and Cahn point out that the configurational entropies of the degenerate states are not the same, and this will then determine the equilibrium state at finite low temperatures. On this basis, the fields of type  $DO_3 \rightarrow B2$  are perhaps better regarded as single phase.

It is particularly noteworthy that the ground state is multiply degenerate in the nearest neighbour model ( $\xi = 0$ ), except for  $x = \frac{1}{2}$ . At compositions near  $x = \frac{1}{4}$ , for example, it is possible for  $z_{AB,1}$  to have its maximum value of 8 (i.e. for there to be no  $B-B$  nearest

neighbour bonds) in a configuration consisting of (i)  $A+B_2$ , (ii)  $DO_3$ , (iii) single phase non-stoichiometric  $B_2$ , or (iv) an almost random structure. When second neighbour interactions are taken into account, this degeneracy disappears and either (i) or (ii) becomes stable even when  $|\xi|$  is very small. This is thus a strong argument for the inclusion of second neighbour interactions in some special situations. There are also degeneracies for  $\xi = \frac{2}{3}$  and  $0.25 \leq x \leq 0.5$ .

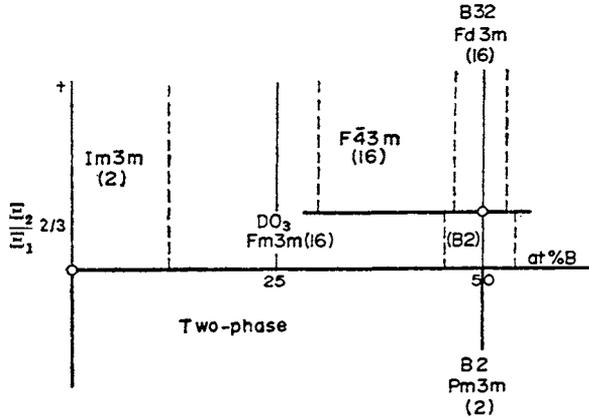


FIG. 6.10. Body-centred cubic ground state diagram showing the state of lowest energy as a function of composition and of the ratio of second to first neighbour interaction energies. Superlattices are identified by Strukturbericht and space-group symbols, with the number of atoms in the unit cell in brackets. Heavy lines show the limits of two-phase fields, and thin lines mark stoichiometric compositions which may be extended into single-phase fields as indicated by the broken lines (after Richards and Cahn, 1971).

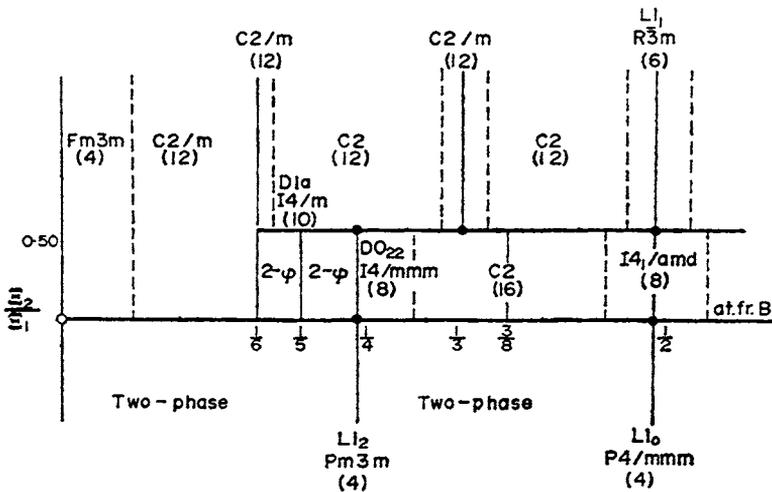


FIG. 6.11. Face-centred cubic ground state diagram for negative  $E_1$ . Notation is similar to Fig. 6.10; the two regions marked 2-φ may be two-phase or polyphase mixtures of  $A_5B$ ,  $A_4B$  and  $A_3B$  superlattices (after Richards and Cahn, 1971).

The ground-state structures for negative  $\xi$  in the more complex f.c.c. diagram (Fig. 6.11) are again two phase, and the  $\text{Cu}_3\text{Au}$  and  $\text{CuAu}$  superlattices are stable. Among the many superlattices predicted for positive  $\xi$  are the known types  $D1a$  ( $\text{Ni}_4\text{Mo}$ ),  $DO_{22}$  ( $\text{Al}_3\text{Ti}$ ) and  $L1_1$  ( $\text{CuPt}$ ), an unknown  $A_5B$  structure, and at  $x \sim 0.5$  an eight-atom cell similar to the  $\text{CuAu II}$  structure (p. 227). Allen and Cahn (1973) have shown that Fig. 6.11 is incorrect in the region  $0 < \zeta < \frac{1}{2}$  and  $\frac{1}{6} < x < \frac{1}{2}$ . The true ground state at  $x = \frac{1}{3}$  is the  $\text{Pt}_2\text{Mo}$  structure (Immm), and for ranges  $\frac{1}{6} - \frac{1}{4}$ ,  $\frac{1}{4} - \frac{1}{3}$  and  $\frac{1}{3} - \frac{1}{2}$  of  $x$  there are 2 or 3-phase, 2-phase and multiphase fields respectively.

For large negative values of  $\xi$  ( $< -\frac{4}{3}$ ) the energy of mixing of the disordered b.c.c. solution is positive even though the ground state is an ordered phase. This also applies to f.c.c. solutions and is contrary to the simple idea that low-temperature ordering is associated with negative  $\Delta_m U$  and phase separation with positive  $\Delta_m U$ . However, the second neighbour interchange energy is unlikely to be larger than the nearest neighbour interchange energy, so the usual association of ordering with the sign of  $\Delta_m U$  is not unreasonable.

Allen and Cahn also investigated the ground state for positive  $\Xi_1$ , and in b.c.c. structures found it to be pure  $A +$  pure  $B$  for all  $\xi > -\frac{2}{3}$  and a mixture of pure  $A$  and the  $B32$  ( $\text{NaTi}$ ) structure for  $\xi < -\frac{2}{3}$ . Similarly for f.c.c. structures,  $A+B$  mixtures are stable for  $\xi > -1$  and mixtures of  $A +$  the  $\text{CuPt}$  superlattice for  $\xi < -1$ . Very large second neighbour interactions are thus necessary to produce superlattices when  $\Xi_1$  is positive.

Structures not included in Figs. 6.10 and 6.11 may also be stable because of higher neighbour interactions or non-central forces. Clapp and Moss (1968) assumed that the observed diffuse X-ray maxima due to short-range ordering (see below) indicate minima of the  $k$ -space potential (p. 113), and they deduced the stable structures for stoichiometric compositions with up to third neighbour real-space interactions. Khachatryan (1962, 1973) uses a similar approach based on the symmetry rules of Landau and Lifshitz (1958) for second-order transitions, and especially on the Lifshitz criterion that minima of the harmonic part of the  $k$ -space potential occur at the special points of the Brillouin zone where symmetry elements intersect. He thus considers only structures generated by composition waves with wave vectors corresponding to the special points, and this excludes some f.c.c. superlattices in Fig. 6.11. On the other hand, Khachatryan includes some b.c.c. structures which are not ground states, at least in a near-neighbour model, and which therefore do not appear in the Richards-Allen-Cahn scheme. An elegant discussion of symmetry rules for order-disorder reactions with a comparison of the various approaches is given by de Fontaine (1975).

The use of second neighbour interactions in calculations of partition functions and entropies is very difficult and for the remainder of this chapter we shall use the nearest neighbour model. We now turn to a discussion of the transition from the disordered to the ordered structure, and we begin by distinguishing between the concepts of long-range and short-range order. In most binary superlattices, the tendency of the like atoms to separate and the unlike atoms to attract each other results in all the  $A$  atoms occupying sites comprising one or more sublattices of the whole structure, and the  $B$  atoms occupying sites which make up different sublattices of the structure. The extent to which this ideal arrangement is achieved is a measure of the long-range order of the assembly. Short-range order is a description of the atomic configuration in the immediate vicinity of an atom; if the average

number of unlike neighbours of an  $A$  atom is higher than would be expected for random distribution, the alloy possesses short-range order. Both concepts may be defined in terms of the probability  $P_{AB}(\mathbf{r})$  introduced in Section 23. The alloy possesses short-range order if  $P_{AB}(\mathbf{r})$  is greater than the random value  $2x(1-x)$  for nearest neighbour values of  $\mathbf{r}$ . (Negative short-range order, or clustering, corresponds to  $P_{AB}(\mathbf{r}) < 2x(1-x)$ ; for convenience, any deviation from the random value is often referred to as short-range order, as discussed in the next section.) The alloy possesses long-range order if  $P_{AB}(\mathbf{r})$  for large  $\mathbf{r}$  approaches a limiting value which is not identical with the random value. It is perfectly possible and usual for there to be appreciable short-range order, but no long-range order.

The distinction between long-range and short-range order is sometimes obscured by the possible existence of antiphase domains. These are regions of perfect or nearly perfect order, at the boundaries of which the roles of two or more of the sets of sites are interchanged. Figure 6.12 shows a simple illustration of the effect in a two-dimensional  $AB$  super-

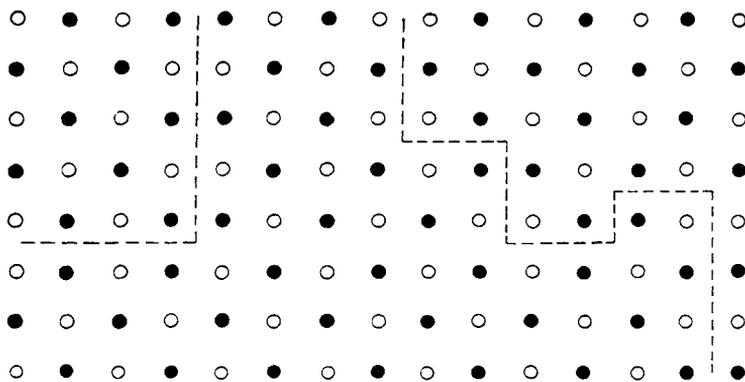


FIG. 6.12. Anti-phase domains in a two-dimensional ordered structure.

lattice. The only atoms not in fully ordered positions are those at the boundaries of the domains, and there is thus a high degree of short-range order. In terms of sublattices which are continuous through all the domains, the long-range order will be small or zero, since there will be nearly equal numbers of the two kinds of atom on each sublattice. Obviously, however, if the domains extend over very many atomic diameters, it would be undesirable to conclude that the structure has no long-range order, and the difficulty can be avoided by considering only a single domain.

Some review articles give a misleading impression that there is no real distinction between a structure of antiphase domains and one possessing short-range order. In fact the domain aggregate may be shown to be thermodynamically unstable, since it involves an increase in energy (at the boundaries) with very little extra entropy to compensate. A structure with genuine short-range order is often the stable configuration because of its high entropy. It is true, of course, that if the size of the antiphase domains continually decreases, the resultant configuration is eventually a state of short-range order only, but at this stage, the domains have lost their individual identity.

A structure of highly ordered antiphase domains is analogous in many ways to a polycrystalline aggregate of strain free grains. The domain boundaries will vanish when true equilibrium is attained, but it is possible in some circumstances to produce a metastable structure in which the domains persist for long periods. From any initial configuration of small domains, the approach to equilibrium occurs in two stages. The domain boundaries first contract, so as to reduce their area, and this is a relatively rapid process, during which the energy decreases continually. It results eventually in an array of planar or nearly planar boundaries, meeting along lines in groups of three at mutual angles of about  $120^\circ$ , and at corners in groups of four. Such an array of either domains or crystal grains is often compared to the structure of a soap froth, and referred to as a foam structure. When it is attained, small displacements of any of the boundaries may lead to an increase in energy.<sup>†</sup> At this stage, the structure changes more slowly, reductions in energy being obtained by movements of boundaries over distances of the order of the domain diameter. As a result of these movements, the smaller domains are progressively eliminated, and the larger domains grow.

There is an important distinction between an array of grains and an array of antiphase domains in a single grain. A grain boundary may separate two crystals of any orientations, but there is a very limited number of different antiphase domains possible with any superlattice structure. Thus in the  $L2_0$  structure, there are only two domains, the corner atoms of the unit cell being occupied by either *A* or *B* atoms. In an  $A_3B$  superlattice with the  $L1_0$  structure, the *A* atoms may occupy any of the four positions of the cubic unit cell, thus giving four different domains.

The foam structure requires the meeting of four different domains at a corner, and hence it is not possible for the  $L2_0$  structure to form such a metastable array. This structure has only one type of antiphase domain boundary, and each such boundary must begin and end on the surface of the grain. The boundaries will straighten rapidly, giving a structure of a few large domains extending right across the crystal. This conclusion that superlattices which have only two types of antiphase domain cannot exist as aggregates of small domains was first pointed out by Bragg (1940).

We have now to formulate our ideas about long- and short-range order in more precise terms, introducing order parameters which are measures of these concepts. In order to evaluate the parameters, we have to use a model, and the nearest neighbour, or Ising, model is the only one which has been found tractable to mathematical analysis. It is useful to note the close analogy between the formation of a phase mixture and the formation of a superlattice, and between the presence of small local aggregates in a solid solution and the presence of short-range order. The analogy is particularly useful for an *AB* superlattice, and all results of order-disorder theory in successive approximations may be derived from the corresponding results of the regular solution theory (for a detailed discussion of this point, see Guggenheim, 1952, pp. 113-16). The simplest assumption in the theory of regular solutions does not consider clustering on an atomic scale. The corresponding approxi-

<sup>†</sup> In a fuller discussion, in Section 35, we show that this condition can be satisfied exactly in two dimensions but not in three dimensions. If the structure arises from randomly nucleated centres, the completely metastable array will not be formed, even in two dimensions.

mation in the theory of order-disorder transformations is similarly a theory of long-range order only. This treatment, which we describe first, is usually called the Bragg-Williams theory. These workers originally developed it in a rather different (but equivalent) manner from that which we shall use (Bragg and Williams, 1934, 1935.)

Consider a crystal of a binary alloy having  $N$  sites, of which  $Nx_A$  sites (the  $A$  sites) may be distinguished in some way from the remaining  $Nx_B$  sites (the  $B$  sites). When the alloy is fully ordered, these two sets of sites are completely occupied by  $A$  atoms and  $B$  atoms respectively. The completely disordered state is taken to be the random distribution, in which there will be  $Nx_A^2$   $A$  atoms and  $Nx_Ax_B$   $B$  atoms on the  $A$  sites, and  $Nx_B^2$   $B$  atoms and  $Nx_Ax_B$   $A$  atoms on the  $B$  sites. We define the partially ordered state by the number of "wrong"  $A$  atoms (i.e.  $A$  atoms on  $B$  sites), which we write  $Nx_w$ ; this is also equal to the number of "wrong"  $B$  atoms. The probability that an  $A$  site is occupied by a "wrong" (i.e.  $B$ ) atom is written  $w_A = x_w/x_A$ , and the probability of its being occupied by a "right" atom is  $r_A = 1 - w_A$ . The corresponding probabilities for the  $B$  sites are  $r_B, w_B$ ; note that  $r_A \neq r_B$  except for an equi-atomic alloy. The definition of long-range order introduced by Bragg and Williams, and now generally used, is

$$L = \frac{r_A - x_A}{1 - x_A} = \frac{r_B - x_B}{1 - x_B} = 1 - \frac{x_w}{x_A x_B}. \quad (26.3)$$

This has a maximum value of  $L = 1$  for  $x_w = 0$ , and  $L = 0$  when  $x_w$  has the random value. For  $x_w > x_A x_B$ ,  $L$  becomes negative, but the long-range order must increase again, since this corresponds only to a different labelling of the sites. For example, in an  $L2_0$  superlattice, the situation when all the  $A$  atoms are on the  $B$  sites ( $L = -1$ ) is physically indistinguishable from the fully ordered state of  $L = +1$ . Hence only the magnitude of  $L$  is significant, and we need only consider  $0 \leq x_w \leq x_A x_B$ .

Let us now consider the equi-atomic superlattice in more detail, confining our attention to the simplest condition (type  $L2_0$ ) in which the atomic sites are situated on two equivalent interpenetrating lattices. Each site has  $z$  nearest neighbours, all of which are situated on the other lattice. As there are  $\frac{1}{2}N$  sites of each type,  $w = 2x_w$  and  $L = 2r - 1$ , where the subscripts for  $w, r$  have been dropped, since  $w_A = w_B$ . In the simplest approximation, we assume that the  $Nr/2$   $A$  atoms and the  $Nw/2$   $B$  atoms are distributed completely randomly amongst the  $A$  positions, and similarly for the  $Nw/2$   $A$  atoms and the  $Nr/2$   $B$  atoms on the  $B$  positions. As on p. 188, this enables us to calculate the number,  $zN_{AB}$ , of nearest neighbour pairs of unlike atoms, and thus to obtain an expression for the free energy. A given atom of the first lattice will have on average  $zr$   $B$  neighbours and  $zw$   $A$  neighbours. The numbers of  $A-A$ ,  $B-B$ , and  $A-B$  pairs are thus  $Nzrw/2$ ,  $Nzrw/2$  and  $Nz(r^2 + w^2)/2$ , and the internal energy may be written

$$\begin{aligned} U &= -Nzr(1-r)(\mathcal{E}_{AA} + \mathcal{E}_{BB}) - Nz\{r^2 + (1-r)^2\}\mathcal{E}_{AB} \\ &= -Nz\mathcal{E}_{AB} - Nzr(1-r)\mathcal{E}. \end{aligned} \quad (26.4)$$

We are assuming random distributions of the atoms on each set of sites, and since any arrangement on the  $A$  sites is independent of the arrangement on the  $B$  sites, the number of

distinguishable ways of arranging the atoms is

$$\frac{(N/2)!}{(Nr/2)!(Nw/2)!}$$

After using Stirling's theorem in the usual manner, we obtain for the configurational free energy

$$G \approx F = -Nz\{\mathcal{E}_{AB} + r(1-r)\mathcal{E}\} + NkT\{r \ln r + (1-r) \ln(1-r)\}. \quad (26.5)$$

To find the equilibrium value of  $r$ , we equate  $\partial G/\partial r = 0$ , and obtain

$$(1/(2r-1)) \ln\{r/(1-r)\} = z\mathcal{E}/kT \quad (26.6)$$

or

$$(2/L) \tanh^{-1} L = z\mathcal{E}/kT. \quad (26.7)$$

Equations (26.6) and (26.7) are identical in form with eqns. (23.7) and (23.9), and they have the same properties. For all temperatures, a possible solution is  $r = \frac{1}{2}$  ( $L = 0$ ); this gives a maximum free energy below a critical temperature  $T_\lambda = -z\mathcal{E}/2k$ , and a minimum free energy above this temperature. The degree of long-range order is thus zero above  $T_\lambda$  and increases with falling temperature below  $T_\lambda$ . In the latter range, there are two further roots of eqn. (26.6) symmetrically disposed about  $r = \frac{1}{2}$ . These roots minimize the free energy, and either of them gives the equilibrium value of  $L$  at a given temperature.

In Fig. 6.13 the equilibrium value of  $L$  is plotted as a function of  $T/T_\lambda$ . The curve is nearly

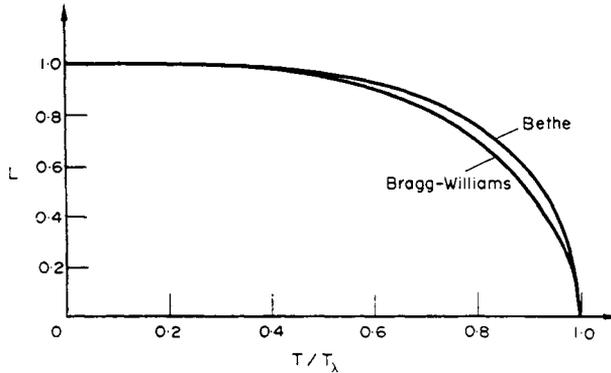


FIG. 6.13. Equilibrium long-range order vs. temperature curve for an  $AB$  superlattice according to the Bragg-Williams and Bethe approximations.

horizontal at first, and then falls more and more steeply as the degree of disorder increases. This behaviour is characteristic of co-operative phenomena, in which the resistance to further disordering (using the term order in a general sense) decreases with decreasing order. The best-known example of such a co-operative phenomenon is the alignment of elementary atomic spins, which leads to ferromagnetism; the Bragg-Williams theory of long-range order is formally analogous to the Weiss theory of ferromagnetism.

The temperature  $T_\lambda$  is often referred to as a  $\lambda$  point in order-disorder and similar theories, the name being taken from the shape of the specific heat anomaly in the neighbourhood of  $T_\lambda$ . From eqn. (26.4), the internal energy may be written

$$U(L) = -Nz\{\bar{\epsilon}_{AB} + (1-L^2)\bar{\epsilon}/4\} \quad (26.8)$$

and the total change in internal energy associated with the ordering process is thus  $-Nz\bar{\epsilon}/4 = NkT_\lambda/2$ . This value is in fair agreement with experiment. The curve of  $U$  as a function of temperature, obtained by combining (26.7) and (26.8), is shown in Fig. 6.14; there is a discontinuity at  $T = T_\lambda$ . The gradient of  $U$  gives the configurational contribution to the specific heat, i.e. the excess specific heat caused by the transformation. This is plotted in Fig. 6.15, together with the experimental curve for the ordering of  $\beta$ -brass, which is an

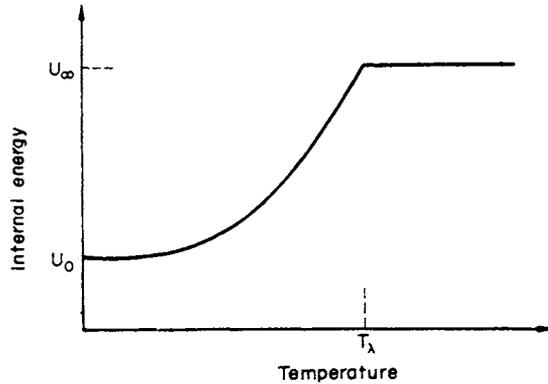


FIG. 6.14. Internal energy of an  $AB$  superlattice according to the Bragg-Williams theory.

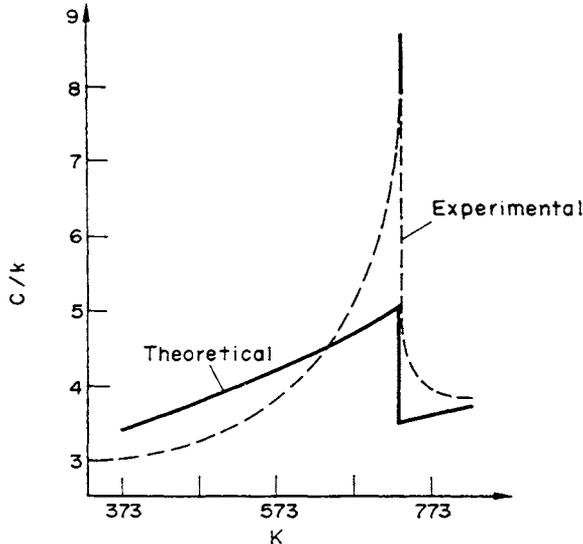
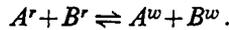


FIG. 6.15. Comparison of the Bragg-Williams prediction of the atomic specific heat with the experimental measurements of Sykes and Wilkinson (1937) on  $\beta$ -brass.

order-disorder transformation of the  $L2_0$ - $A2$  type. In this approximation, the specific heat remains finite at all temperatures, and the ordering reaction is not a first-order phase change.

The original treatment of this theory of long-range order did not refer specifically to the nearest neighbour model, and preceded Bethe's development of this model for alloys. Bragg and Williams considered the change in energy  $W$  associated with the interchange of an  $A$  and a  $B$  atom, both of which were originally in "right" positions, and, finally, in "wrong" positions. Use of Boltzmann's equation then gives an expression for  $r_A/w_A$  in terms of  $W$ ; alternatively this expression may be derived by considering the kinetic balance of the rates at which right atoms interchange to give wrong atoms, and vice versa. These two processes may be written symbolically as an equation (analogous to a chemical equation)



Note that interchanges of right  $A$  atoms with wrong  $B$  atoms, or vice versa, are without significance, since the right or wrong status of neither atom is affected by the interchange.

The rate at which the forward reaction takes place is proportional to the number of right  $A$  atoms,  $N(x_A - x_w)$ , and to the number of right  $B$  atoms,  $N(x_B - x_w)$ , whilst the rate of the back reaction is proportional to  $(Nx_w)^2$ . At equilibrium, these two rates may be equated, and

$$(x_A - x_w)(x_B - x_w)/x_w^2 = k_w/k_r, \quad (26.9)$$

where  $k_w$ ,  $k_r$  are the rate constants for the forward and back reactions respectively. The equilibrium constant may then be written  $k_w/k_r = \exp(-W/kT)$ .<sup>†</sup> On substituting from eqn. (26.3), the following expression is obtained for the long-range order parameter

$$L = 1 - \frac{[4x_A x_B \{\exp(W/kT) - 1\} + 1]^{1/2} - 1}{2x_A x_B \{\exp(W/kT) - 1\}}. \quad (26.10)$$

The energy  $W$  introduced into this theory is not a constant, but depends upon  $L$ . Clearly  $W$  is zero in a state of complete disorder, since there is then no energy change on interchanging any  $A$  and  $B$  atoms,<sup>‡</sup> and has a maximum value  $W_0$  in the completely ordered state. Bragg and Williams made the simple assumption that there is a linear relation  $W = W_0 L$ . For the equi-atomic superlattice, eqn. (26.10) simplifies to

$$L = 1 - \frac{2 \exp(W/2kT) - 1}{\exp(W/kT) - 1} = \tanh(W/4kT). \quad (26.11)$$

Comparing this with eqn. (26.7), we see that the expressions are identical if  $W = W_0 L = -2z\bar{E}L$ . It may readily be seen that this equivalence applies to the more general expression (26.10) and the corresponding equation derived by the zeroth approximation of the quasi-chemical method. The two theories are in fact identical if the substitution  $W_0 = -2z\bar{E}$  is

<sup>†</sup> Note that this derivation shows the implicit assumption is being made that the entropy of activation is equal for the two processes. For a further discussion, see Part II, Chapter 18.

<sup>‡</sup> More correctly, the energy change is equally likely to be positive or negative.

made, and the Bragg-Williams assumption that  $W = W_0L$  is equivalent to the quasi-chemical assumption of random mixing in the two sets of positions. This serves to emphasise that although the zeroth approximation is developed in terms of the nearest neighbour model, it makes no real use of the short-range character of the binding forces.

We next consider the nature of the changes introduced by removing the restrictive approximation that the atoms are arranged randomly on the two sets of sites. Clearly, it is possible to develop a treatment of order-disorder phenomena based on the assumption of independent nearest neighbour interaction, and this will be exactly analogous to the first approximation of the quasi-chemical method. Such a theory of order-disorder reactions was first given by Bethe (1935). Bethe's analysis was more complex than that given above, but the equivalence of the two methods was proved by Rushbrooke (1938) and Fowler and Guggenheim (1940). We shall not repeat the derivations, but simply assume that we can transcribe the appropriate formulae from Section 24. In particular, we find for the critical temperature of long-range order in an  $L2_0$  superlattice

$$T_\lambda = -\bar{\mathcal{E}}/k \ln(z/(z-2)). \quad (26.12)$$

The equilibrium roots of  $r$  below  $T = T_\lambda$  are given by

$$\frac{1 - [r/(1-r)]}{[r/(1-r)]^{1/z} - [r/(1-r)]^{(z-1)/z}} = \exp(-\bar{\mathcal{E}}/kT), \quad (26.13)$$

which is the equivalent of (24.16). In Fig. 6.13 we compare the curve for  $r$ , and hence for  $L = 2r - 1$ , given by eqn. (26.13) for an  $L2_0$  superlattice with that given by eqn. (26.6). Both equations predict that the long-range order falls to zero at a critical temperature ( $\lambda$  point), but the Bethe theory shows that the long-range order decreases more slowly than is predicted by the Bragg-Williams approximation. It follows from the general correspondence of the methods, that the order-disorder curves of Fig. 6.13 are the same as the solubility limit curves of Fig. 6.9.

The configurational internal energy will be given by

$$U = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - 2\bar{N}_{AB}\bar{\mathcal{E}}/N) \quad (26.14)$$

(see p. 191), where  $\bar{N}_{AB}$  is a function of temperature and is determined by eqn. (24.7) with  $N_A = N_B = \frac{1}{2}N$ . As  $T \rightarrow 0$  K, the superlattice structure becomes perfect, and  $\bar{N}_{AB} \rightarrow \frac{1}{2}N$ . Thus at 0 K

$$U_0 = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - \bar{\mathcal{E}}). \quad (26.15)$$

At very high temperatures,  $\bar{N}_{AB}$  is equal to the value  $N_{AB}^0$  corresponding to  $r = \frac{1}{2}$  and random mixing on each of the sublattices, i.e. to  $N/4$ . We then have

$$U_\infty = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - \frac{1}{2}\bar{\mathcal{E}}), \quad (26.16)$$

and the total energy of disordering is  $-Nz\bar{\mathcal{E}}/4$ , as in the zeroth approximation. The present treatment shows, however, that only part of the configurational energy of ordering disappears below the  $\lambda$  point. From eqns. (26.14), (26.15), and (26.16)

$$(U_T - U_0)/(U_\infty - U_0) = 2(1 - 2\bar{N}_{AB}/N). \quad (26.17)$$

This equation gives implicitly the value of the configurational energy  $U_T$  as a function of  $T$  or of the order parameter  $L$ . At any temperature,  $\overline{N_{AB}}$  may be found from the following equation, which is analogous to eqn. (24.7),

$$\xi^2 = (r - \xi)(1 - r - \xi) \exp(2\mathcal{E}/kT),$$

where  $2\xi = 1 - 2\overline{N_{AB}}/N$ . The solution may be written in the form (24.8) as

$$\xi = 2\xi/(\beta + 1) = 2r(1 - r)/(\beta + 1),$$

and  $\beta$  is given by eqn. (24.9) with  $\mathcal{E}$  replaced by  $-\mathcal{E}$  and  $x$  replaced by  $r$ . The value of  $r$  in these equations is given by (26.6). At  $T = T_\lambda$  and higher temperatures,  $r = \frac{1}{2}$  and  $\beta = \exp(-\mathcal{E}/kT)$ . This gives

$$\overline{N_{AB}} = \frac{N \exp(-\mathcal{E}/kT)}{2[1 + \exp(-\mathcal{E}/kT)]}. \quad (26.18)$$

Above the critical point, the curve of  $U_T$  against  $T$  is represented by the equation

$$U_T - U_0 = \frac{-Nz\mathcal{E}}{2[1 + \exp(-\mathcal{E}/kT)]}. \quad (26.19)$$

Below the critical point, it is not possible to write a simple analytical expression for  $U_T - U_0$ . The curve of  $U_T - U_0$  is plotted against  $T/T_\lambda$  in Fig. 6.16, together with the corres-

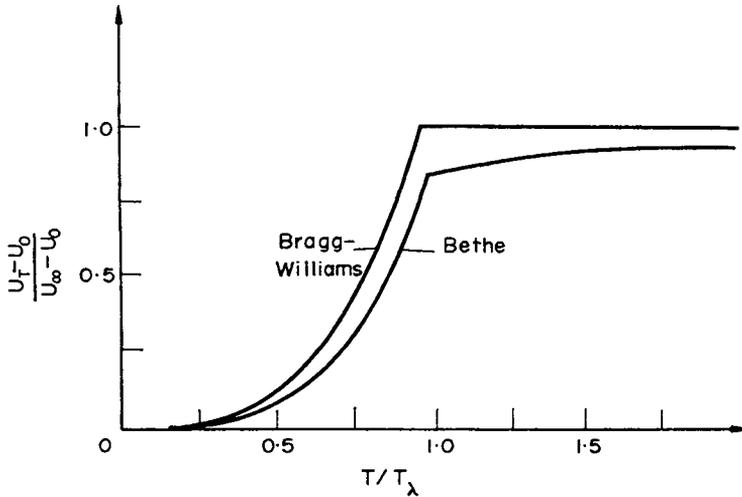


FIG. 6.16. Comparison of the internal energy vs. temperature curves given by the zeroth (Bragg-Williams) and first (Bethe) approximations for an  $AB$  superlattice.

ponding curve for the zeroth approximation. We may also calculate from (26.19) that

$$(U_\lambda - U_0)/(U_\infty - U_0) = (z - 2)/(z - 1).$$

For a b.c.c. structure, sixth-sevenths of the total ordering energy is destroyed below the critical temperature according to this approximation, and the remaining one-seventh dis-

appears more gradually as the temperature is raised above  $T_\lambda$ . The gradient of  $U_T$  gives the configurational specific heat, i.e. the excess specific heat above the Debye curve. This is shown in Fig. 6.17.

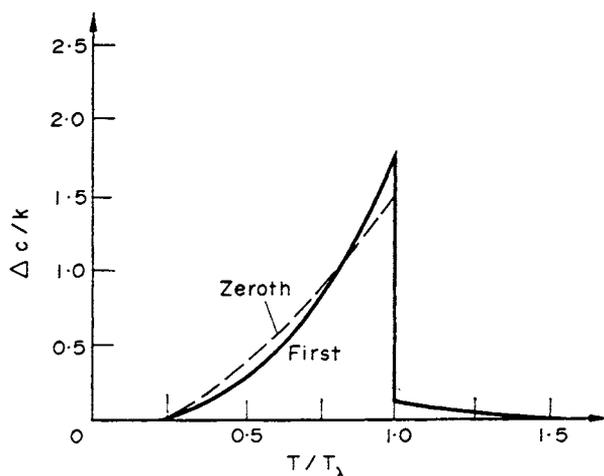


FIG. 6.17. Comparison of the excess atomic specific heat curves given by the zeroth and first approximations.

The principal advantage of the Bethe treatment over the Bragg-Williams approximation is in this prediction of an excess specific heat above the disordering temperature, since this is always observed experimentally. The physical interpretation of the effect is in terms of the concept of short-range order. Although the long-range order becomes zero at the critical temperature, the number of unlike nearest neighbour pairs remains larger than would be found in a random distribution, so that short-range order persists above the critical point. The destruction of long-range order has been frequently compared to the melting of a solid; local crystallinity persists in the liquid, and local order in the disordered solution.

Bethe introduced a short-range order parameter  $\sigma$ , which is defined so that  $\sigma = 1$  for perfect long-range order, and  $\sigma = 0$  for a completely random distribution. In these respects,  $\sigma$  is equivalent to  $L$ , but the definition is directly in terms of the nearest neighbours of an atom rather than of the segregation into different lattices. The general definition of the short-range order parameter is

$$\sigma = (\overline{N_{AB}} - N_{AB}^0) / (N_{AB}^0 - N_{AB}^1), \quad (26.20)$$

where  $\overline{N_{AB}}$ ,  $N_{AB}^0$  are the actual value of  $N_{AB}$  and the value for a purely random arrangement, as before, and  $N_{AB}^1$  is the value of  $N_{AB}$  for the fully ordered state. In this simple case we have been considering,  $N_{AB}^0 = N/2$ , and  $N_{AB}^1 = N/4$ , so that

$$\sigma = (4\overline{N_{AB}}/N) - 1. \quad (26.21)$$

Comparing this with eqn. (26.17), we see

$$(U_T - U_0)/(U_\infty - U_0) = 1 - \sigma. \quad (26.22)$$

With this definition of ordering, the internal energy of ordering is proportional to the degree of short-range order. At  $T_\lambda$ , for example, the short-range order parameter  $\sigma = \frac{1}{7}$ .

Although Bethe's theory is successful in explaining short-range order, detailed comparison of experimental and theoretical results shows rather poor agreement. Attempts at more exact theories usually take the form of higher cluster approximations or series expansions (see Takagi, 1941). We shall not consider these here, but it is interesting to refer briefly again to the exact solutions obtained in two dimensions. A square lattice forms a superlattice of the type described above, there being no  $AA$  or  $BB$  pairs in the fully ordered state. The rather surprising result, first obtained by Onsager (1944) and discussed more fully by Wannier (1945), is that the specific heat of the structure has a logarithmic infinity at  $T = T_\lambda$ . The heat content remains continuous, i.e. there is no latent heat, and the change is thermodynamically of the second order (see pp. 226-8). This result is at variance with the predictions of all the approximate calculations, according to which there is a finite discontinuity in the specific heat. Moreover, in the exact two-dimensional solution, the curve is symmetrical about  $T = T_\lambda$ , so that as much of the energy of disordering is required above  $T_\lambda$  as below it. Exact solutions in three dimensions are unlikely to be obtained; it is probable that such a solution would still contain a logarithmic infinity in the specific heat, but would not be symmetrical about  $T_\lambda$ .

The above discussion has been confined to the simplest equi-atomic superlattice. The next most important example is the structure of type  $L1_2$ , formed in  $A_3B$  alloys. We shall not develop the theory of this superlattice, but shall merely describe the results. We now have  $w_B = 3w_A = 4x_w$ , and the Bragg-Williams definition of long-range order becomes

$$L = (4r_A - 1)/3 = 1 - 16x_w/3. \quad (26.23)$$

Attempts to apply the first approximation, or Bethe theory, to the  $L1_2$  superlattice result in a contradiction; the hypothesis of the non-interference of nearest neighbour pairs does not lead to equilibrium long-range order at all. The reason for this difficulty seems to be that at least four atom sites are needed to define a unit cell of the superlattice. It is thus necessary to use a higher approximation of the quasi-chemical theory (cluster variation method), embodying the hypothesis of the non-interference of tetrahedral groups of atoms. The stability of the  $L1_2$  structure at the  $A_3B$  composition was first shown in this way by Yang (1945), and the treatment was considerably extended in later papers by Yang and Li (1947) and Li (1949). These papers also considered the more difficult problems associated with non-stoichiometric compositions. When  $x_B$  is a variable, the use of the quasi-chemical method ensures that all properties are symmetrical about  $x_B = \frac{1}{2}$ , so, in this treatment, superlattices based on both  $A_3B$  and  $AB_3$  necessarily appear in the same binary alloys.

The results of the Bragg-Williams approximation and of the tetrahedral cluster method are shown in Fig. 6.18, which gives the equilibrium degree of long-range order as a function of temperature. In contrast to the  $L2_0$  type of superlattice,  $L$  drops discontinuously to zero at the critical temperature. The transformation thus requires a latent heat, and is correctly

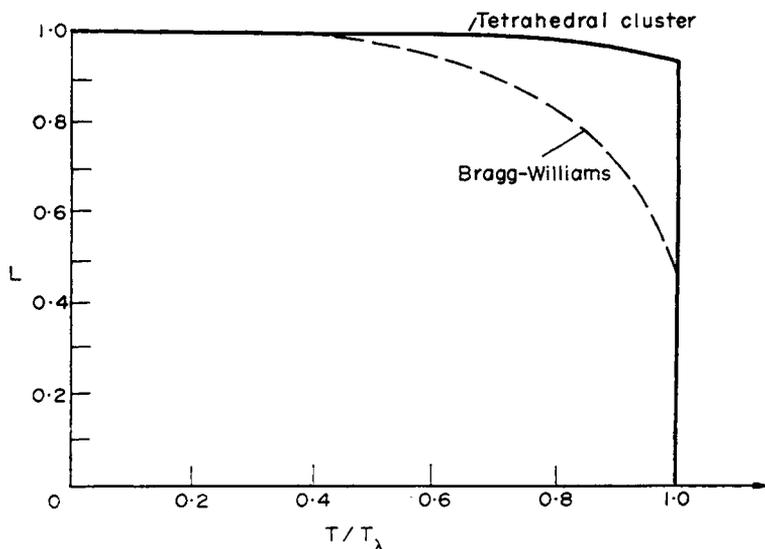


FIG. 6.18. Equilibrium long-range order vs. temperature curve for an  $A_3B$  superlattice of type  $L1_2$  according to the Bragg-Williams and tetrahedral cluster approximations (after Guggenheim, 1952).

described as a thermodynamic phase change of the first order. The transformation temperature, at which the superlattice phase with a finite degree of long-range order is in equilibrium with the solution of zero long-range order, is given by

$$\left. \begin{aligned} T &= -z\bar{E}/7.3k && \text{(Bragg-Williams),} \\ T &= -z\bar{E}/14.6k && \text{(tetrahedral cluster).} \end{aligned} \right\} \quad (26.24)$$

These values are obtained by numerical solution of the equations, no simple analytical expressions being obtained.

An approximate theory of order in alloys due to Cowley (1950 a, b) differs in some respects from the cluster variation methods. Cowley considers a set of short-range order coefficients which effectively define the probability  $P_{AB}(\mathbf{r})$  for each of the different shells (successively larger values of  $r$ ) around an atom. The internal energy is defined in terms of the interactions of pairs of atoms in different shells, so that the theory is not confined to nearest neighbour interactions. It is assumed as an approximation that the order coefficients are independent of each other, this corresponding approximately to the independence of pairs in the first approximation of the quasi-chemical theory.

Cowley's theory has the advantage of being related rather more closely to the experimental X-ray measurements of the  $P_{AB}(\mathbf{r})$  quantities than are the above descriptions. The results of work on the diffuse X-ray scattering from solid solutions are frequently expressed in terms of the order coefficients used by Cowley (positive or negative for short-range order and clustering; see Section 27), together with size coefficients representing the displacements of the atoms from the ideal sites of a structure with the measured lattice parameter. This work will not be described here; reviews have been given by Averbach (1956) and

Sivertsen and Nicholson (1961). From the mean pair probabilities for  $\text{Cu}_3\text{Au}$ , Cowley deduced that the values of  $\bar{E}$  for the first three shells (sets of nearest neighbours) are given by  $\bar{E}_1/k = 358^\circ$ ,  $\bar{E}_2/k = -34^\circ$ ,  $\bar{E}_3/k = -19^\circ$ . For the  $L2_0$  superlattice, Cowley's theory gives the same results as the Bragg-Williams theory; the disordering temperature for the other types of superlattice differs slightly from the predictions of the above theories.

The question of whether superlattice changes are first- or second-order thermodynamic transformations has attracted much discussion. The approximate solutions of the nearest neighbour model suggest that some transitions involve a continuous decrease of the long-range order parameter to zero (second-order change), whilst others predict a discontinuous fall at  $T = T_c$  (first-order change). Thus changes from  $L2_0$  or  $D0_3$  structures to disordered b.c.c. structures may be second order, whilst most others are first order. In these second-order changes, both superlattice and disordered phases are non-close-packed structures, in which the nearest neighbours of any one atom are not nearest neighbours of each other. In his review article on order-disorder phenomena in metals, Guttman (1956) considerably clarified these difficult problems, and the following description is taken mainly from his original work.

The thermodynamic classification of transitions is based on the order of the lowest derivative of  $G$  which shows a discontinuity at the transition temperature. If the values of this function are  $G^\alpha$  and  $G^\beta$  for the two forms concerned, the transition temperature is defined by the condition  $G^\alpha = G^\beta$ . In a normal first-order transition, there are discontinuities in the derivatives of  $G$  with respect to temperature (entropy) and with respect to pressure (volume). In a second-order transition, the entropy and volume are continuous, but their derivatives (e.g. the specific heat and the compressibility) are not. It follows that the enthalpy is also continuous, and the transition has an anomalous specific heat but no latent heat.

A well-known difficulty in the theory of even-order transitions arises if the difference in the free energies of the two phases  $\Delta G^{\alpha\beta}$  is expanded as a Taylor series in terms of the difference in temperature  $\Delta T = T - T^{\alpha\beta}$ . This gives

$$\Delta G^{\alpha\beta} = -(\Delta S^{\alpha\beta}) \Delta T + \frac{1}{2}(\partial^2 G / \partial T^2) \Delta T^2 + \frac{1}{6}(\partial^3 G / \partial T^3) \Delta T^3 + \dots$$

The sign of  $\Delta G^{\alpha\beta}$  is determined by the sign of the first non-vanishing term on the right, and this is the  $n$ th term for an  $n$ th-order transition. For a first- or third-order transition,  $\Delta G^{\alpha\beta}$  changes sign with  $\Delta T$ , but for an even-order transition, it has the same sign above and below  $T^{\alpha\beta}$ . Thus if attempts are made to draw  $G^\alpha$  and  $G^\beta$  curves, they do not intersect at the transition temperature, but only touch, with one curve always below the other. This apparently results in either  $\alpha$  or  $\beta$  always being stable.

This difficulty is thought to arise because of the attempt to extrapolate properties of the assembly into regions in which they do not exist. We cannot really draw separate  $G^\alpha$  and  $G^\beta$  curves for a second-order superlattice transition, but only a single curve which represents an ordered structure below the transition temperature and a substantially disordered structure above it. The existence of the singularity in  $\partial^2 G / \partial T^2$  renders the analytic expansion above invalid.

A second-order transition in a solid solution implies not only a vanishing of the latent

heat of transition, but also a continuity of composition. The transition temperature is marked by a single line on the phase diagram, and the coexistence of ordered and disordered phases is not possible at equilibrium. This provides one of the most convenient experimental tests of the order of a transition, it being very difficult in practice to distinguish between latent heats and anomalous specific heats. In the early stages of order-disorder theory, there was a tendency to regard the disordering of almost all superlattices as second-order transitions, but it has become increasingly evident that many of these changes are first order. It seems obvious that the transition must be first order when there is a change in lattice symmetry, as in the  $L1_0$  (tetragonal) superlattice formed from the f.c.c. structure, and the most convincing evidence of the existence of equilibrium two phase regions comes from alloys undergoing this structural change.†

Rhines and Newkirk (1953) have suggested that all superlattice transitions are first-order phase changes, and they presented some evidence, based on electrical resistivity measurements, to show that this is true even for the  $L2_0$ - $A2$  transition in copper-zinc alloys. In general, this suggestion has not been substantiated by later workers; for example, a two-phase region was sought but not found in a very careful investigation of copper-zinc by Beck and Smith (1952). Moreover, measurements of long-range order by X-ray methods tend to support the division of superlattice changes into first- and second-order types. For the  $L2_0$  structures in copper-zinc and silver-zinc alloys, the long-range order decreases continually with temperature (Chipman and Warren, 1950; Muldower, 1951), the decrease becoming more and more rapid as the critical temperature is approached. Whilst the experimental techniques do not allow the definite conclusion that there is no discontinuity in  $L$ , it is certain that the magnitude of any discontinuous drop is much smaller than that for the transitions from  $L1_0$  or  $L1_2$  superlattices in copper-gold and copper-platinum alloys. In these alloys, the amount of long-range order decreases only slightly as the temperature is increased within the superlattice region, and then drops discontinuously to zero (Keating and Warren, 1951; Roberts, 1954; Walker, 1952). Thus these measurements tend to confirm the qualitative predictions of the nearest neighbour model for all kinds of superlattice.

We have assumed throughout this section that the ordering tendency is provided by the chemical interactions between nearest neighbour atoms. Suggestions are often made that an atomic size disparity is an important factor in lowering the energy of the superlattice relative to that of a random arrangement, a view first proposed by Hume-Rothery and Powell (1935). It is not entirely clear whether this distinction is meaningful (see p. 201), but it is obvious that a strain energy calculation of the type used in Section 25 has no relevance to the ordering energy, which must arise from more localized interactions. At the same time, there are a number of superlattices known to have large unit cells, and the formation of these structures cannot be explained on the basis of purely nearest neighbour interactions.

The best-known example of a superlattice with a large unit cell is the structure CuAu II. At low temperatures, equi-atomic copper-gold alloys form a tetragonal superlattice of the  $L1_0$  form (CuAu I) and, at high temperatures, the structure is disordered f.c.c. However,

† The Bragg-Williams theory predicts a second-order transition, but the first approximation of the quasi-chemical theory correctly predicts a first-order transition (Guggenheim, 1952).

it was found in 1936 that in an intermediate temperature range, now known to be  $\sim 380^\circ\text{C}$  to  $\sim 405^\circ\text{C}$ , the equilibrium structure has a more complex unit cell with dimensions  $10a$ ,  $a$ , and  $c$ , where  $c$  and  $a$  are the tetragonal parameters (Johansson and Linde, 1936). The structure may be considered to be formed by introducing antiphase domain boundaries parallel to (100) planes at every five lattice planes of the tetragonal structure, so that the repeat unit has to contain two opposite "domains". The complete diffraction effects, which are rather complex, also indicate a small expansion across the antiphase domain boundaries. Direct evidence for the existence of these domains has been obtained by transmission electron-microscopy of thin films (Pashley and Presland 1958–9; Glossop and Pashley, 1959). The structure consists of regions within which the antiphase domain boundaries are parallel planes; two kinds of region are observed, corresponding to the two perpendicular  $a$  directions of the tetragonal cell.

Other examples of structures with large unit cells are now known, and are formed by a similar disturbance of the structure of a simple superlattice, but the spacing of the antiphase boundaries is not always constant, sometimes varying with temperature and composition (Schubert *et al.*, 1954, 1955). From this point of view, it seems attractive to regard the structures as superlattices containing some kind of fault. Nevertheless, and in spite of our description of the CuAu II structure as a tetragonal superlattice with regularly spaced antiphase domain boundaries, it is quite clear that, in these alloys at least, the large superlattice is the thermodynamically stable form. This cannot be explained by the nearest neighbour model because an  $A$  atom has the same number of nearest neighbour  $B$  atoms in both the CuAu I and the CuAu II structures if the slight deviations from cubic symmetry are ignored.

The thermodynamics of the transitions CuAu I  $\rightarrow$  CuAu II  $\rightarrow$  disordered f.c.c. have been investigated by Oriani and Murphy (1958). The latent heats for the two changes are 888 and 1590 J g atom $^{-1}$  respectively; the former value seems remarkably large in view of the slight structural rearrangement involved. The results show other disagreements with the quasi-chemical theory. For example, the total disordering energy per g-atom for the transition CuAu I  $\rightarrow$  disordered f.c.c. can be evaluated, and the temperature  $T_1$  for this transition in the absence of the CuAu II structure can be estimated. The ratio of the disordering energy to  $RT_1$  is found to be 0.73 compared with 0.5 of the zeroth approximation and 1.37 given by Li's treatment of the  $L1_2-A1$  change using the tetrahedral cluster method. The heats of formation of both ordered and disordered phases were measured by Oriani and Murphy, and give independent estimates of  $\bar{E}$ . These values are 1570 and 1940 J g atom $^{-1}$  respectively. It thus seems clear that the quasi-chemical treatment is extremely unsatisfactory for the CuAu superlattices.

## 27. FLUCTUATIONS IN SOLID SOLUTIONS: SHORT-RANGE ORDER AND CLUSTERING

We have emphasized that the pair probability functions will generally not have the values corresponding to a completely random arrangement, even in a substantially disordered solid solution. These functions, or related order parameters, may be measured by X-ray methods, and give quantitative information on the amount of short-range order or

clustering. In general, short-range order, for which  $P_{AB}(\mathbf{r}) > 2x(1-x)$  when  $\mathbf{r}$  represents a nearest neighbour vector, is expected if the solid solution undergoes a superlattice transition at lower temperatures. Clustering (negative short-range order), for which  $P_{AB}(\mathbf{r}) < 2x(1-x)$  for nearest neighbour vectors, is expected if there is a solubility gap at lower temperatures.<sup>†</sup>

The X-ray measurements of short-range order give the average number of neighbours of given kinds possessed by any atom. We now examine the nature of the qualification expressed by the word "average". An instantaneous picture of a solid solution would obviously show that many atoms had a surrounding configuration quite different from the average. This is because the average configuration is obtained by a dynamic averaging process over all the available configurations. In postulating this, we are assuming some mechanism for changing from one configuration to another; this is achieved in practice by atomic diffusion.

For kinetic applications, we require a quantitative estimate of how the arrangement of a small region of the assembly may momentarily differ from the average arrangement. This is provided by the theory of fluctuations, and we express the probable state of the assembly in terms of the root mean square deviation from the average arrangement. We begin by considering a solid solution in which the atomic arrangement is completely random. The statistical properties of such a solution are especially easy to derive, since all arrangements have equal probability.

We consider in all  $N$  atoms, each having  $z$  nearest neighbours. For each atom, we fix attention first on a particular neighbour. The probability that this neighbour is a  $B$  atom is  $x$ , and hence there will be  $Nx$  of the neighbours we are considering which are  $B$  atoms. Now consider another neighbour of each of our  $N$  atoms. Since the probabilities are independent, there will again be a chance  $x$  that each such neighbour is a  $B$  atom. The number of  $B-B$  pairs among the  $N$  sets of two neighbours will thus be proportional to  $Nx^2$ , whereas the number of  $A-B$  pairs will be  $Nx(1-x)$ . Similarly, the number of  $A-A$  pairs and of  $A-B$  pairs will be proportional to  $N(1-x)^2$  and  $Nx(1-x)$ . We thus have that for  $z = 2$ , the numbers of atoms having two  $A$  atoms, an  $A$  atom and a  $B$  atom, and two  $B$  atoms as nearest neighbours are  $N(1-x)^2$ ,  $2Nx(1-x)$  and  $Nx^2$  respectively. Exactly similar reasoning shows that for  $z = 3$ , the numbers of atoms with three  $A$  atoms, two  $A$  atoms and one  $B$  atom, one  $A$  atom and two  $B$  atoms, and three  $B$  atoms are proportional to  $N(1-x)^3$ ,  $3N(1-x)^2x$ ,  $3N(1-x)x^2$  and  $Nx^3$  respectively. The relative numbers of neighbours of the different types are given by the terms of the expansions  $\{(1-x)+x\}^2$  and  $\{(1-x)+x\}^3$ . We conclude by extension of this reasoning that the relative numbers of atoms which have as nearest neighbours  $zA$  atoms,  $(z-1)A$  atoms and one  $B$  atom,  $(z-2)A$  atoms and two  $B$  atoms, . . . , one  $A$  atom and  $(z-1)B$  atoms, and  $zB$  atoms are given by the terms of the binomial expansion

$$\{(1-x)+x\}^z.$$

<sup>†</sup> For many years, gold-nickel alloys were quoted as an extreme example of the failure of the quasi-chemical theory, since they have a solubility gap at low temperatures, and X-ray measurements (Flinn, *et al.*, 1953) indicated positive short-range order at high temperatures. Later work by Munster and Sagel (1959) has shown that the interpretation of the earlier X-ray results may have been incorrect, and there may actually be clustering above the solubility gap. It is possible that this is a necessary result, and is independent of the assumptions of any particular model.

This expression is not a continuous mathematical function, i.e. there is no chance of finding an atom with a fractional number of  $A$  atoms among its nearest neighbours. There is a distinction between the average number of neighbours of a given type and the most probable number. Thus if  $x = 0.7$  and  $z = 4$ , each atom has on the average 1.2  $A$  atoms and 2.8  $B$  atoms as nearest neighbours. The most probable neighbours for any atom are one  $A$  atom and three  $B$  atoms.

The above result is not confined to nearest neighbour probabilities, since we have simply calculated the probability of having a given number of  $B$  atoms amongst the  $z$  atoms which happen to be neighbours of any atom. We are often interested in rather larger clusters of (say)  $n$  atoms. The probabilities of our finding 0, 1, 2, 3, ...,  $n-1$ ,  $n$   $B$  atoms in such a cluster when the average composition is  $x$  are given by the terms of the expansion  $\{(1-x)+x\}^n$ , and the general term, i.e. the probability of the group containing (say)  $m$   $B$  atoms is

$$\binom{n}{m} (1-x)^{n-m} x^m. \quad (27.1)$$

This formula may be used as a first approximation in the problem of nucleus formation in a solid solution.

If the solution is very dilute, the above law may be simplified. We let  $x \rightarrow 0$ , increasing  $n$  at the same time, so that the average number of solute atoms in the group,  $nx$ , remains constant. The general term of the binomial expansion can now be written as

$$\frac{n!}{m!(n-m)!} \left(\frac{nx}{n}\right)^m \left(1 - \frac{nx}{n}\right)^{n-m} = \frac{(nx)^m}{m!} \left(1 - \frac{nx}{n}\right)^n \left[ \frac{n!}{(n-m)! n^m (1 - nx/n)^m} \right].$$

By using Stirling's theorem, the term in square brackets reduces to

$$\left(1 - \frac{nx}{n}\right)^{-m} \left(1 - \frac{m}{n}\right)^{-n+m-\frac{1}{2}} \exp(-m)$$

and as  $n \rightarrow \infty$  with  $nx$  constant, this may be replaced by unity. At the same time, the term  $(1 - nx/n)^n$  tends to  $\exp(-nx)$ . We thus have, finally, for the probability of finding  $mB$  atoms amongst  $n$  atoms:

$$\{(nx)^m/m!\} \exp(-nx). \quad (27.2)$$

This is Poisson's distribution. The expression has a maximum value when  $m = nx$ , i.e. when the number of solute atoms in the group equals the average number, and the concentration fluctuation is zero. The fluctuation is most conveniently expressed by the root mean square deviation from the average value; this is the conventional mean deviation of statistical theory.

Since  $(m - nx)^2 = m^2 - (nx)^2 + 2nx(nx - m)$ , and the mean deviation of  $(nx - m)$  is zero, we have

$$\overline{(m - nx)^2} = \sum_{m=0}^n [m^2 - (nx)^2] \{(nx)^m/m!\} \exp(-nx).$$

The first part of this expression is

$$\sum_{m=0}^n (m^2/m!) (nx)^m \exp(-nx) = nx(1+nx),$$

so that  $\overline{(m-nx)^2} = nx$ , and the mean deviation is  $(nx)^{1/2}$ . Expressed as a composition fluctuation, we write  $m = n(x + \Delta x)$ , and the mean deviation is given by

$$\overline{(\Delta x)^2} = x/n. \quad (27.3)$$

The concentration fluctuations thus increase as the number of atoms in the group diminishes. Provided  $n$  is large enough (or, equivalently,  $x$  is small enough), to justify the use of Poisson's equation, the mean deviation is proportional to  $n^{-1/2}$ .

A similar method may be used to find the mean deviation of the more exact binomial expansion. We then obtain

$$\text{or } \left. \begin{aligned} \overline{(m-nx)^2} &= nx(1-x) \\ \overline{(\Delta x)^2} &= x(1-x)/n. \end{aligned} \right\} \quad (27.4)$$

Obviously, this reduces to (27.3) when  $x$  is small.

In the above discussion, we have not been concerned with the effects of temperature at all. This is because all arrangements of the ideal solution have the same energy. Temperature has no effect on the magnitude of the fluctuations, although it will determine the rate of fluctuation. In a real solution, we must consider the change in free energy which is associated with a fluctuation. The probability of the fluctuation is then related to this change in energy by the Boltzmann factor.

Consider a region of solid solution which has initially a concentration of  $B$  atoms equal to the average or equilibrium concentration,  $x$ . As a result of a fluctuation, one half of this region may be supposed to contain a fraction  $x + \Delta x$  of  $B$  atoms, whilst the other half contains a fraction  $x - \Delta x$  of  $B$  atoms. If the free energy per atom is  $g(x)$ , we have

$$\Delta g = \frac{1}{2} \{g(x + \Delta x) + g(x - \Delta x)\} - g(x).$$

And after expanding in a Taylor series, we obtain

$$\Delta g = \frac{1}{2} (\partial^2 g / \partial x^2) (\Delta x)^2$$

provided that  $\Delta x$  is small. If the original volume element contained  $n$  atoms, the total change in the free energy of the assembly is  $n \Delta g$ . Since the average value of  $n \Delta g$  must be  $\frac{1}{2} kT$ , the mean square fluctuation is

$$\overline{(\Delta x)^2} = kT/n(\partial^2 g / \partial x^2) \quad (27.5)$$

and this equation replaces (27.4) when the energy of the assembly is a function of the arrangement. The value of  $\partial^2 g / \partial x^2$  for the zeroth approximation of the quasi-chemical theory is derived from eqn. (23.5). Substituting this into (27.5),

$$\overline{(\Delta x)^2} = \left( \frac{n}{x(1-x)} - \frac{2z\bar{\epsilon}}{kT} n \right)^{-1}. \quad (27.6)$$

This equation reduces to (27.4) when  $\mathcal{E} \rightarrow 0$ , this being the condition for the solution to be ideal. A similar result is obtained when  $T$  becomes very large, so that, at high temperatures, the fluctuations approach those expected in a random solid solution, as is physically obvious. At lower temperatures, the fluctuations depend on the sign of  $\mathcal{E}$ . If  $\mathcal{E}$  is positive, there is a tendency to clustering, and the fluctuations are greater than those in an ideal solution. As the temperature is lowered, the fluctuations become larger, and according to (27.5) are infinite when  $\partial^2 g / \partial x^2 = 0$ . This equation is, of course, only valid for small  $x$ , but it is clear that large fluctuations may be expected under these conditions since there is no first-order dependence of the free energy on the atomic arrangement. Thus when the spinodal is reached, all concentrations have the same probability. When  $\mathcal{E}$  is negative (tendency to ordering) the fluctuations are always smaller than would be found in an ideal solution.

The above treatment, though giving a better approximation to the fluctuations in a solution than do the purely statistical expressions, is logically rather unsatisfactory. We have used the zeroth approximation for the free energy, and so have assumed that the equilibrium configuration is completely random, even though the energy depends on this configuration. In a more accurate treatment,  $\partial^2 g / \partial x^2$  would be calculated from the first approximation, and used in eqn. (27.5). Unfortunately, this does not lead to any simple analytical expression. We should note, however, the difference between the value of  $(\overline{\Delta x})^2$  given by eqn. (27.6) and the mean number of neighbours of a given kind possessed by any atom. Thus above the critical temperature, the equilibrium, or mean, composition in any region of  $n$  atoms

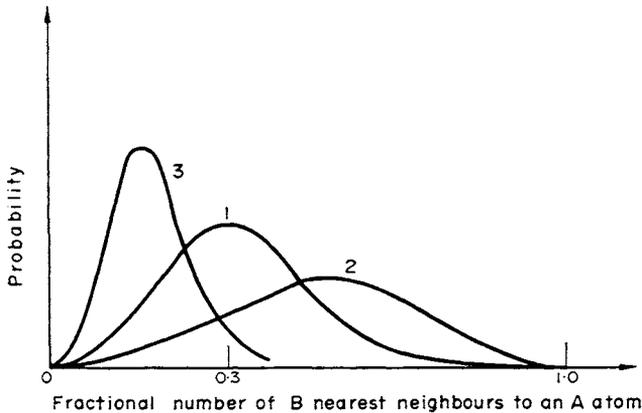


FIG. 6.19. Probability that a given fraction of the nearest neighbours of an  $A$  atom are  $B$  atoms in a solid solution for which  $x = 0.3$  (after Smoluchowski, 1951). Curve 1: Ideal solution, Curve 2: Short-range order. Curve 3: Clustering.

is always  $x$  so long as  $n$  is large. For smaller values of  $n$ , the concentration fluctuations become correspondingly large, and (27.4) ceases to apply. We have already seen that for  $\mathcal{E}$  negative, the tendency to order persists above the critical temperature, and the number of  $A-B$  pairs is greater than the random value. Exactly the same reasoning holds for  $\mathcal{E}$  positive; we expect a finite configurational specific heat corresponding to the atomic clusters which gradually disappear with increasing temperature. This clustering or short-range

ordering is quite distinct from the fluctuations described above, and expresses the fact that if we view on a sufficiently fine scale, the average distribution is no longer uniform. For  $\bar{\epsilon}$  positive we have a short-range phase mixture, just as we have short-range order for  $\bar{\epsilon}$  negative. Thermodynamically, the solution remains one phase, since thermodynamic concepts only have meaning for large numbers of atoms. Figure 6.19 shows qualitatively the number of nearest neighbours of type  $A-B$  for  $x = 0.3$ .

## REFERENCES

- ALLEN, S. M. and CAHN, J. W. (1972) *Acta Metall.* **20**, 423; (1973) *Scripta Metall.* **7**, 1261.  
 AVERBACH, B. L. (1956) *Theory of Alloy Phases*, p. 301, American Society for Metals, Cleveland, Ohio.  
 BECK, L. H. and SMITH, C. S. (1952) *Trans. Am. Inst. Min (Metall.) Engrs.* **194**, 1079.  
 BETHE, H. (1935) *Proc. R. Soc. A*, **150**, 552.  
 BRAGG, W. L. (1940) *Proc. Phys. Soc.* **52**, 105.  
 BRAGG, W. L. and WILLIAMS, E. J. (1934) *Proc. R. Soc. A*, **145**, 699; (1935) *Ibid.* **151**, 540.  
 BROWN, W. B. and LONGUET-HIGGINS, H. C. (1951) *Proc. R. Soc. A*, **209**, 416.  
 CAHN, J. W. (1961) *Acta Metall.* **9**, 795; (1969) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 1, Institute of Metals, London.  
 CAHN, J. W. and HILLIARD, J. E. (1958) *J. Chem. Phys.* **28**, 258; (1971) *Acta Metall.* **19**, 151.  
 CHIPMAN, W. B. and WARREN, B. E. (1950) *J. Appl. Phys.* **21**, 696.  
 CLAPP, P. C. and MOSS, S. C. (1968) *Phys. Rev.* **171**, 754.  
 COTTRELL, A. H. (1955) *Theoretical Structural Metallurgy*, Arnold, London.  
 COWLEY, J. M. (1950a) *J. Appl. Phys.* **21**, 24; (1950b) *Phys. Rev.* **77**, 664.  
 DE FONTAINE, D. (1975) *Acta Metall.* **23**, 553.  
 ESHELBY, J. D. (1954) *J. Appl. Phys.* **25**, 255; (1956) *Solid State Phys.* **3**, 79.  
 FLINN, P. A., AVERBACH, B. L., and COHEN, M. (1953) *Acta Metall.* **1**, 664.  
 FOWLER, R. H. and GUGGENHEIM, E. A. (1940) *Proc. R. Soc. A*, **174**, 189.  
 FREEDMAN, J. F. and NOWICK, A. S. (1958) *Acta Metall.* **6**, 176.  
 FREUNDLICH, H. (1926) *Colloid and Capillary Chemistry* (transl. of 3rd German edition).  
 FRIEDEL, J. (1955) *Phil. Mag.*, **46**, 514.  
 GIBBS, J. W. (1878) *Trans. Conn. Academy* **3**, 439; *Collected Works*, I, p. 219, Yale University Press, New Haven, Connecticut.  
 GLOSSOP, A. B. and PASHLEY, D. W. (1959) *Proc. Roy. Soc. A*, **250**, 132.  
 GUGGENHEIM, E. A. (1948) *Trans. Faraday Soc.* **44**, 1007; (1952) *Mixtures*, Clarendon Press, Oxford.  
 GUINIER, A. (1959) *Solid State Phys.* **9**, 293.  
 GUTTMAN, L. (1956) *Solid State Phys.* **3**, 146.  
 HARDY, A. H. (1953) *Acta Metall.* **1**, 202.  
 HILLERT, M. (1970) *Phase Transformations*, p. 181, American Society for Metals, Metals Park, Ohio.  
 HUME-ROTHERY, W. and POWELL, H. M. (1935) *Z. Kristallogr.* **91**, 23.  
 HUME-ROTHERY, W. and RAYNOR, G. V. (1954) *The Structure of Metals and Alloys*, Institute of Metals, London.  
 JOHANNSON, C. H. and LINDE, J. O. (1936) *Ann. Phys.* **25**, 1.  
 KAUFMAN, L. (1967) *Phase Stability of Metals and Alloys*, p. 125, McGraw-Hill, New York; (1969) *Prog. Mater. Sci.* **14**, 57.  
 KAUFMAN, L. and BERNSTEIN, H. (1970) *Computer Calculation of Phase Diagrams*, Academic Press, New York.  
 KEATING, D. T. and WARREN, B. E. (1951) *J. Appl. Phys.* **22**, 286.  
 KHACHATURYAN, A. G. (1962) *Phys. Met. and Metallg.* **13**, 493; (1973) *Phys. Stat. Sol.* **60**, 9.  
 KIRKWOOD, J. G. (1938) *J. Chem. Phys.* **6**, 70.  
 KURATA, M., KIKUCHI, R., and WATARI, T. (1953) *J. Chem. Phys.* **21**, 434.  
 LANDAU, L. D. and LIFSHITZ, E. M. (1958) *Statistical Physics*, Addison-Wesley, Reading, Mass.  
 LAWSON, A. W. (1947) *J. Chem. Phys.* **15**, 831.  
 LI, Y. Y. (1949) *Phys. Rev.* **76**, 972.  
 LONGUET-HIGGINS, H. C. (1951) *Proc. R. Soc. A*, **205**, 247.  
 MOTT, N. F. and NABARRO, F. R. N. (1940) *Proc. Phys. Soc.* **52**, 86.  
 MULDAWER, L. (1951) *J. Appl. Phys.* **22**, 663.

- MUNSTER, A. and SAGEL, K. (1959) *The Physical Chemistry of Metallic Solutions and Intermetallic Compounds*, paper 2D, HMSO, London.
- ONSAGER, L. (1944) *Phys. Rev.* **65**, 117.
- ORIANI, R. A. (1956) *Acta Metall.* **4**, 15; (1959) *The Physical Chemistry of Metallic Solutions and Intermetallic Compounds*, paper 2A, HMSO, London.
- ORIANI, R. A. and MURPHY, W. K. (1958) *J. Phys. Chem. Solids* **6**, 277.
- PASHLEY, D. W. and PRESLAND, A. E. B. (1958-9) *J. Inst. Metals* **87**, 419.
- PRIGOGINE, I., BELLEMANS, A., and MATHOT, V. (1957) *The Molecular Theory of Solutions*, North-Holland, Amsterdam.
- RHINES, F. N. and NEWKIRK, J. B. (1953) *Trans. Am. Soc. Metals* **45**, 1029.
- RICHARDS, M. J. and CAHN, J. W. (1971) *Acta Metall.* **19**, 1263.
- ROBERTS, B. W. (1954) *Acta Metall.* **2**, 597.
- RUSHBROOKE, G. S. (1938) *Proc. R. Soc. A*, **166**, 296.
- SCHUBERT, K., KIEFER, B., and WILKENS, M. (1954) *Z. Naturforsch.* **9a**, 987.
- SCHUBERT, K., KIEFER, B., WILKENS, M., and HAUFLER, R. (1955) *Z. Metallk.* **46**, 692.
- SHIMOJI, M. and NIWA, K. (1957) *Acta Metall.* **5**, 496.
- SIVERTSEN, J. M. and NICHOLSON, M. E. (1961) *Prog. Mater. Sci.* **9**, 303.
- SMOLUCHOWSKI, R. (1951) *Phase Transformation in Solids*, p. 149, Wiley, New York.
- SWAN, E. and URQUHART, A. R. (1927) *J. Phys. Chem.* **21**, 251.
- SYKES, C. and WILKINSON, H. (1937) *J. Inst. Metals* **61**, 223.
- TAKAGI, Y. (1941) *Proc. Phys. Math. Soc. Japan* **23**, 44.
- THOMSON, J. J. (1888) *Application of Dynamics to Physics and Chemistry*.
- TILLER, W. A., POUND, G. M., and HIRTH, J. P. (1970) *Acta Metall.* **18**, 225.
- TRIVEDI, R. (1970) *Metall. Trans.* **1**, 921.
- WALKER, C. B. (1952) *J. Appl. Phys.* **23**, 118.
- WANNIER, G. H. (1945) *Rev. Mod. Phys.* **17**, 50.
- YANG, C. N. (1945) *J. Chem. Phys.* **13**, 66.
- YANG, C. N., and LI, Y. Y. (1947) *Chin. J. Phys.* **7**, 59.
- ZENER, C. (1951) *J. Appl. Phys.* **22**, 372.

## CHAPTER 6

### *Solid Solutions*

#### 22. PAIR PROBABILITY FUNCTIONS: THERMODYNAMIC PROPERTIES

A solid phase containing two or more kinds of atom, the relative proportions of which may be varied within limits, is described as a solid solution. Terminal solid solutions are based on the structures of the component metals; intermediate solid solutions may have structures which are different from any of those of the constituents. Most solid solutions are of the substitutional type, in which the different atoms are distributed over one or more sets of common sites, and may interchange positions on the sites. In interstitial solutions, the solute atoms occupy sites in the spaces between the positions of the atoms of the solvent metal; this can only happen when the solute atoms are much smaller than the atoms of the solvent.

We must also distinguish between ordered and disordered solid solutions. In the fully ordered state each set of atoms occupies one set of positions, so that the atomic arrangement is similar to that of a compound. This is only possible at compositions where the ratios of the numbers of atoms of different kinds are small integral numbers, but the atomic arrangement may still be predominantly ordered in this way for alloys of arbitrary composition. In disordered solid solutions, the atoms are distributed among the sites they occupy in a nearly random manner. This classification is only approximate, and we shall formulate these concepts more precisely.

The definition of the unit cell, and the concept of the translational periodicity of the lattice, lose their strict validity when applied to a disordered solid solution. The mean positions of the atoms, considered as mathematical points, will no longer be specified exactly by (5.8), since there will be local distortions depending on the details of the local configurations. Moreover, a knowledge of the type of atom at one end of a given interatomic vector no longer implies knowledge of the atom at the other end, as it does for a pure component or a fully ordered structure. In a solid solution, precise statements of this nature have to be replaced by statements in terms of the probability of the atom being of a certain type.

For many purposes, the strict non-periodicity of the structure is not important, since most physical properties are averages over reasonably large numbers of atoms. Thus the positions of X-ray diffraction maxima depend only on the average unit cell dimensions,

and their intensities depend only on the mean concentrations of atoms of different kinds and the mean interatomic distances. An approximate description in which the structure is regarded as having a unit cell of fixed size, with atomic positions occupied by identical scattering centres of averaged atomic scattering factor, thus suffices for a description of the main features of the X-ray diffraction pattern. A lattice vector of this structure may actually connect two unlike atoms, but is regarded as connecting two average atoms.

If two metals have the same crystal structure, they may form a single solid solution, and the lattice parameter then varies continuously with composition from the value characteristic of one pure metal to that of the other. In some assemblies, the edge length of the unit cell, or the volume per atom, is approximately linear with the atomic fraction of solute; this is described as Vegard's law. The work of Hume-Rothery and his collaborators (see Hume-Rothery and Raynor, 1954) has shown that a necessary (but not sufficient) condition for the formation of extensive terminal solutions is that the sizes of the atoms in the pure components shall not differ by more than  $\sim 15\%$ . This is called the size factor rule.

The localized distortion around a solute atom cannot be observed easily, but it may be studied by X-ray diffraction techniques in which the diffuse background scattering is measured. This type of X-ray measurement also enables us to deduce the probability of a given interatomic vector connecting atoms of given types. In theoretical treatments of the properties of solid solutions, it is usual to treat all atoms as situated on the sites obtained from the mean lattice, and this approximation is adequate in view of the severe limitations of the theories in other respects. The way in which the sites are occupied may then be described by a set of probability factors.

Consider a solid solution containing atomic fractions  $x_A, x_B, x_C, \dots$ , of components  $A, B, C, \dots$ , all of which may occupy any of the atomic sites. Then the probability that a particular site is occupied by a particular atom is just equal to the atomic fraction of that kind of atom, whatever the nature of the atomic forces. The probability that a particular vector joins two atoms of specified types, however, is dependent on the details of the distribution, and hence on the interatomic forces. Consider a vector  $\mathbf{r}$  linking two atomic positions. There will be many such vectors in a crystal; suppose that in a group  $N(\mathbf{r})$  of them, there are  $N_{AB}(\mathbf{r})$  which connect together an  $A$  atom and a  $B$  atom. Then we can define the probability

$$P_{AB}(\mathbf{r}) = \lim_{N(\mathbf{r}) \rightarrow \infty} \frac{N_{AB}(\mathbf{r})}{N(\mathbf{r})}. \quad (22.1)$$

This is sometimes called a pair density function, by analogy with similar functions used in the descriptions of non-crystalline materials.

Obviously we may define a pair probability function for each possible kind of occupancy of the two sites,  $A-A, A-C, B-B, B-C$ , etc. However, the probabilities so defined are not completely independent, since their values must be consistent with the overall composition of the solid solution. Consider, for example,  $N(\mathbf{r})$  vectors  $\mathbf{r}$  in a binary assembly. They will join together  $2N(\mathbf{r})$  atoms, of which  $2x_A N(\mathbf{r})$  must be of type  $A$ . The vectors linking  $A-A$  atom pairs will contribute  $2N_{AA}(\mathbf{r})$   $A$  atoms to this total, and those linking

$A-B$  pairs will contribute  $N_{AB}(\mathbf{r})$   $A$  atoms. Hence

$$\left. \begin{aligned} 2N_{AA}(\mathbf{r}) + N_{AB}(\mathbf{r}) &= 2x_A N(\mathbf{r}) \\ \text{or } P_{AB}(\mathbf{r}) &= 2\{x_A - P_{AA}(\mathbf{r})\}. \end{aligned} \right\} \quad (22.2)$$

Similarly,

$$P_{AB}(\mathbf{r}) = 2\{x_B - P_{BB}(\mathbf{r})\}.$$

Only one of the three probabilities is independent in this case; we shall take it to be  $P_{AB}(\mathbf{r})$ . There are equivalent relations reducing the number of independent probabilities for ternary and higher assemblies, but we shall only consider binary assemblies in detail. In such an assembly,  $x_A + x_B = 1$ , and it is useful to employ a single composition variable  $x = x_B$ ,  $(1-x) = x_A$ . The composition of the alloy then changes from pure  $A$  to pure  $B$  as  $x$  increases from 0 to 1. Occasionally, however, it is preferable to retain the  $x_A, x_B$  notation in order to show whether complex relations are symmetrical with regard to the components. We shall also need the notation  $N_A = N(1-x)$ ,  $N_B = Nx$  for the numbers of  $A$  and  $B$  atoms in a crystal of  $N$  atoms.

Suppose we have an assembly in which we have values of  $P(\mathbf{r})$  for all possible vectors  $\mathbf{r}$  and all possible combinations of atoms. The question then arises whether or not this is a complete statistical description of the occupancy of the sites. In general, it is not, and we should also write down the probabilities of groups of three, four, five, or more sites being occupied by atom groups of specified types. Such probability factors are introduced in some theoretical treatments, but we shall not have occasion to consider them in detail. The pair probability functions may be deduced from X-ray measurements, as already mentioned, but no experimental method is known for determining triplet or higher probability functions.

A pair probability function must possess the point group symmetry of the Bravais lattice; that is,  $P(\mathbf{r})$  and  $P(\mathbf{r}')$  must be equal if  $\mathbf{r}$  and  $\mathbf{r}'$  are vectors which can be transformed into each other by a rotation or reflection symmetry operation. The behaviour of  $P(\mathbf{r})$  as  $\mathbf{r}$  becomes large falls into one of two categories. In a solid solution which is described as disordered,  $P(\mathbf{r})$  in the limit of large  $\mathbf{r}$  approaches the value it would have for a completely random distribution of atoms over the available sites. This is the situation we shall discuss in most of this chapter. In a structure described as possessing long-range order, the limit of  $P(\mathbf{r})$  as  $\mathbf{r} \rightarrow \infty$  is different from the random value; this is discussed further in Section 26.

In determining the form of an alloy phase diagram, we are interested in the variation of the free energy of a phase with composition. This means that at fixed temperature and pressure, we need to know the heat and entropy of mixing, and various derived functions. The thermodynamic quantities are related to the parameters which specify the distribution of the atoms over the available sites, but any attempt at precise calculation of the energies of different distributions is formidably difficult. Thus, when analytic expressions are required it is necessary to use a simple model. Much of the theoretical work on solid solutions is restricted to binary solutions formed from two metals of the same crystal structure and nearly equal atomic diameters. Under these conditions, the drastic approximations which are used to obtain the required energy expressions are reasonably valid.

We shall assume throughout that the entropy of a solid solution may be separated into configurational and thermal parts. The thermal or vibrational entropy is obtained from the heat motion of the atoms; the configurational entropy results from the randomness of location of the atoms on their sites, and may be defined as the extra entropy possessed by a hypothetical solid solution with all atoms occupying point sites over the corresponding entropy of the pure components. This separation of the entropy is equivalent to a factorization of the partition function.

In the simplest model (the "ideal solution"), the thermal partition function is determined only by the number of atoms of each kind present, and the configurational partition function corresponds to a completely random distribution of the atoms. The formation of an alloy from its constituents then involves no change in thermal entropy, and the entropy of mixing is equal to the configurational entropy of a random arrangement. The assumption of zero thermal entropy of mixing, though unjustified, is often made in more complex treatments, which attempt to calculate the configurational partition function when the arrangement is not random. The model used in most calculations of this type is that of central force nearest neighbour interactions, and the approach is known as the quasi-chemical theory. There are also statistical models in which volume changes and changes in thermal entropy are not neglected, but other simplifying assumptions have to be made in order to obtain useful results. A completely different model focuses attention on the effects of differences in the sizes of the atoms, and calculates the energy term by treating each solute atom as a centre of dilatation in an elastic medium.

These various models are discussed in some detail in the remaining sections of this chapter. None of them is very satisfactory, and at present it seems that experimental measurements of thermodynamic functions are much more accurate than theoretical predictions. In the remainder of this section we examine the important thermodynamic properties of solutions and their relation to the equilibrium diagram.

The thermodynamic properties of a single component are completely determined by the external constraints but have to be measured from an arbitrary zero; that is, a standard reference state must be specified. This is also true of a solution, and it is often convenient to use the pure components under the same temperature and pressure as the reference state. Thus suppose that  $g_A^{\alpha 0}$ ,  $g_B^{\alpha 0}$  are the free energies per atom of the pure components in the  $\alpha$  phase, each measured from an arbitrary zero.† (The  $\alpha$  phase may, of course, be metastable in either or both components.) The free energy per atom of a homogeneous  $\alpha$  solution of composition  $x$  may then be expressed relative to the same zeros as

$$g^\alpha = (1-x)g_A^{\alpha 0} + xg_B^{\alpha 0} + \Delta_m g^\alpha, \quad (22.3)$$

where  $\Delta_m g^\alpha$  is called the free energy of mixing or the free energy of formation of the  $\alpha$  solution, and is measured from the standard state of the unmixed components.

Suppose Fig. 6.1 represents the curve of  $g^\alpha$  against  $x$  at fixed temperature and pressure. An alloy of composition  $x_4$  has a free energy per atom represented by point  $P$  if it exists

† The zeros may be chosen so that at some particular temperature and pressure  $g_A^{\alpha 0} = g_B^{\alpha 0}$ , but this is not then true for other temperatures and pressures.

as a homogeneous  $\alpha$  solid solution, and by point  $Q$  if it is a mixture of pure  $A$  and pure  $B$  (each in the  $\alpha$  form). If the assembly contains two  $\alpha$  solutions of different composition, the  $g^\alpha$  curve gives the free energy per atom in each, and hence the mean free energy per atom. Thus the mean free energy per atom of an alloy of composition  $x_4$  is represented by point  $R$  if it exists as a two-phase mixture of  $\alpha$  solutions of compositions  $x_5$  and  $x_6$ . Clearly, the relative stability of the single solid solution and the phase mixture depends on whether  $R$  is above or below  $P$ . It is readily seen that if the  $g^\alpha$  curve is concave upwards for all  $x$  (i.e. if  $\partial^2 g^\alpha / \partial x^2 > 0$  for all  $x$ ), the single solid solution is always more stable than a phase mixture

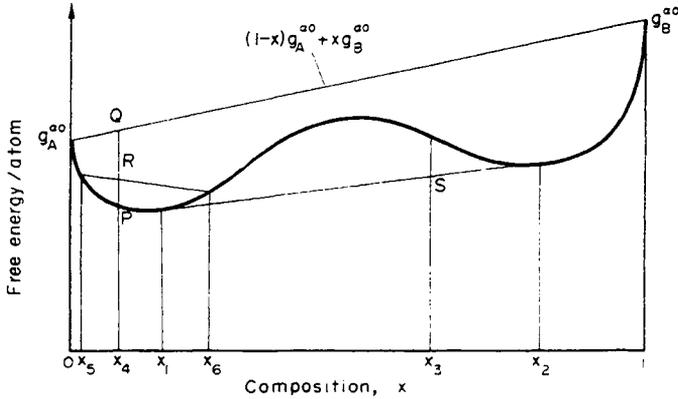


FIG. 6.1. Schematic free energy vs. composition curve for a solid solution.

of two  $\alpha$  solutions. However, if the  $g^\alpha$  curve contains a region of negative  $\partial^2 g^\alpha / \partial x^2$ , as shown in Fig. 6.1, the equilibrium condition will be a phase mixture for certain compositions. For the alloy of composition  $x_3$ , the lowest free energy is represented by the point  $S$ , and is obtained from a mixture of  $\alpha$  phases of compositions  $x_1$  and  $x_2$ . These compositions are those corresponding to the points of contact of the common tangent to the two parts of the  $g^\alpha$  curve. All other alloys in which  $x_1 < x < x_2$  also exist in equilibrium as a mixture of  $\alpha$  phases of compositions  $x_1$  and  $x_2$ , the relative proportions of the phases being fixed by the overall composition.

From this discussion, we see that the limits of solubility  $x_1$  and  $x_2$  are defined by the condition

$$(\partial g^\alpha / \partial x)_{x_1} = (\partial g^\alpha / \partial x)_{x_2} = \{g^\alpha(x_2) - g^\alpha(x_1)\} / (x_2 - x_1). \tag{22.4}$$

Since the  $g^\alpha$  curve is essentially equivalent to the  $\Delta_m g^\alpha$  curve tilted about the straight line  $(1-x)g_A^{\alpha 0} + xg_B^{\alpha 0}$ , this condition may also be written

$$(\partial \Delta_m g^\alpha / \partial x)_{x_1} = (\partial \Delta_m g^\alpha / \partial x)_{x_2}. \tag{22.5}$$

So far as the equilibrium properties are concerned, it is immaterial whether we consider the variation of  $g^\alpha$  or  $\Delta_m g^\alpha$  with  $x$ .

The chemical potential (partial free energy per atom) of  $A$  in the  $\alpha$ -phase is defined by the equation

$$g_A^\alpha = \partial G / \partial N_A = g_A^{\alpha 0} + \partial(N \Delta_m g^\alpha) / \partial N_A. \tag{22.6}$$

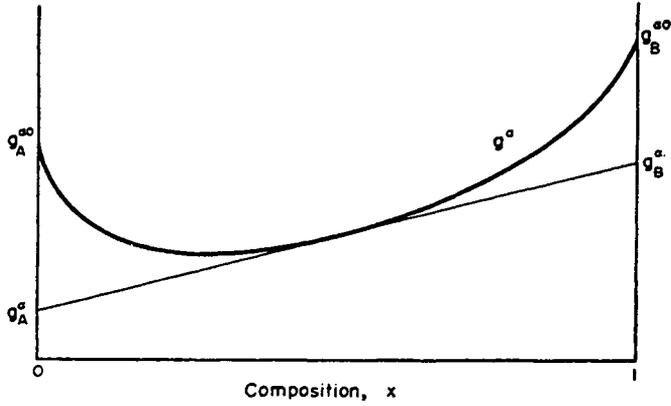


FIG. 6.2. Relation of chemical potentials to free energy vs. composition curve.

The mean free energy per atom is related to the chemical potentials by the equation

and

$$\left. \begin{aligned} g^{\alpha} &= (1-x)g_A^{\alpha} + xg_B^{\alpha} \\ \partial g^{\alpha} / \partial x &= g_B^{\alpha} - g_A^{\alpha} \end{aligned} \right\} \quad (22.7)$$

These relations are illustrated in Fig. 6.2. The tangent to  $g^{\alpha}$  at the composition  $x$  intercepts the  $x = 0$  ordinate to give the chemical potential  $g_A^{\alpha}$  and the  $x = 1$  line to give the chemical potential  $g_B^{\alpha}$ . Clearly, for the common tangent of Fig. 6.1,

$$g_A^{\alpha}(x_1) = g_A^{\beta}(x_2), \quad g_B^{\alpha}(x_1) = g_B^{\beta}(x_2). \quad (22.8)$$

This is the condition for phase equilibrium in terms of the appropriate intensive thermodynamic quantities (chemical potentials), corresponding to the condition of a minimum value of the extensive quantity  $G$ .

We now consider the possibility of two different phases,  $\alpha$  and  $\beta$ , existing in the assembly. These phases may represent different states of matter, or may be solid solutions of diffe-

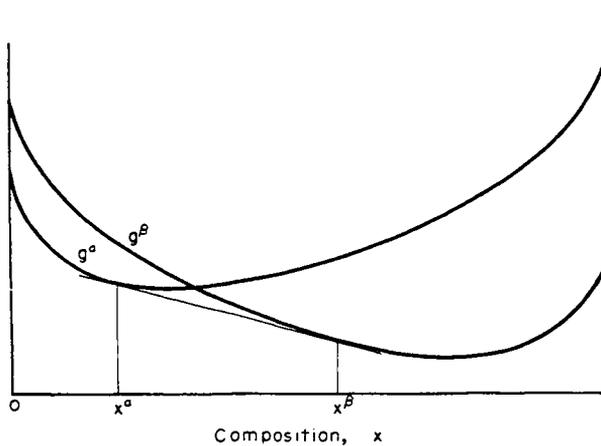


FIG. 6.3. Free energy vs. composition curves for phases  $\alpha$  and  $\beta$ .

rent crystal structure. The free energy of the  $\beta$  phase will be

$$g^\beta = (1-x)g_A^{\beta 0} + xg_B^{\beta 0} + \Delta_m g^\beta, \quad (22.9)$$

where  $g_A^{\beta 0}$ ,  $g_B^{\beta 0}$  are measured from the same zeros as  $g_A^{\alpha 0}$ ,  $g_B^{\alpha 0}$  respectively. Thus  $g_A^{\beta 0} - g_A^{\alpha 0}$  is the free energy per atom required to change pure  $A$  from the  $\alpha$  phase to the  $\beta$  phase. A possible configuration of the free energy curves  $g^\alpha$  and  $g^\beta$  is shown in Fig. 6.3. It follows from the previous discussion that the equilibrium state of the assembly is homogeneous  $\alpha$  phase for  $x < x^\alpha$ , homogeneous  $\beta$  phase for  $x > x^\beta$  and a mixture of  $\alpha$  and  $\beta$  phases for  $x^\alpha < x < x^\beta$ . The solubilities or solvus lines  $x^\alpha$ ,  $x^\beta$  are determined by the relation

$$\left(\frac{\partial g^\alpha}{\partial x}\right)_{x^\alpha} = \left(\frac{\partial g^\beta}{\partial x}\right)_{x^\beta} = \frac{g^\beta(x^\beta) - g^\alpha(x^\alpha)}{x^\beta - x^\alpha}. \quad (22.10)$$

It should be noted that eqn. (22.10) is not equivalent to  $(\partial \Delta_m g^\alpha / \partial x) = (\partial \Delta_m g^\beta / \partial x)$  because of the different reference states used for  $\Delta_m g^\alpha$  and  $\Delta_m g^\beta$ . It is sometimes convenient to rewrite eqns. (22.3) and (22.9) as

$$\left. \begin{aligned} g^\alpha &= (1-x)g_A^{\alpha 0} + xg_B^{\alpha 0} + \Delta_n g^\alpha, \\ g^\beta &= (1-x)g_A^{\beta 0} + xg_B^{\beta 0} + \Delta_n g^\beta, \end{aligned} \right\} \quad (22.11)$$

where  $\Delta_n g^\alpha$ ,  $\Delta_n g^\beta$  are both referred to pure  $A$  and pure  $B$  in their equilibrium states. Equation (22.10) is then equivalent to

$$(\partial \Delta_n g^\alpha / \partial x)_{x^\alpha} = (\partial \Delta_n g^\beta / \partial x)_{x^\beta}. \quad (22.12)$$

Free-energy-composition curves of the type shown in Fig. 6.3 are typical of those expected in a simple eutectic assembly in which the components have different crystal structures  $\alpha$  and  $\beta$ ; those shown in Fig. 6.1 are typical of those from a eutectic assembly in which both metals have the same crystal structure. At temperatures below the eutectic temperature, the liquid free-energy curve lies entirely above the common tangent. Above the eutectic temperature, this curve intersects the common tangent, as shown in Fig. 6.4. Alterna-

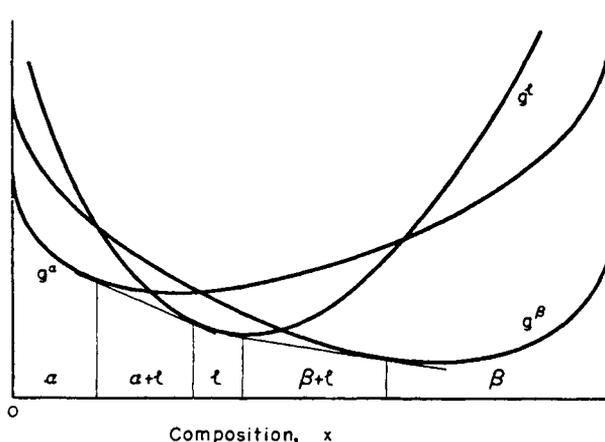


FIG. 6.4. Free energy vs. composition curves for a eutectic reaction.

tively, if the liquid free-energy curve first crosses the  $g^{\alpha}$  curve at  $x = 0$ , or the  $g^{\beta}$  curve at  $x = 1$ , a peritectic reaction will result. The same principles apply when further (intermediate) phases are considered; these may be solid solutions having relatively flat free-energy curves, or compounds of fixed composition, at which the free energy has a very sharp minimum. The relation of free-energy-composition curves to the equilibrium diagram is discussed in many textbooks; see, for example, Cottrell (1955) for an elementary account. It is now possible to store the basic thermodynamic information for a given system in a computer and to use a program which will enable any required information on phase equilibria or derived thermodynamic functions to be obtained. The necessary free-energy functions are generally obtained from a mixture of experimental measurements, empirical expansions, and information derived from measured phase diagrams. Of particular value for work on phase transformations is the ability in this way to obtain reliable estimates of metastable phase relations in otherwise inaccessible regions of the phase diagram; the usual theory of pearlite formation, for example, requires knowledge of the metastable austenite-ferrite and austenite-cementite equilibria below the eutectoid temperature. This kind of work was pioneered by Kaufman (1967, 1969) and has been surveyed by Kaufman and Bernstein (1970) and Hillert (1970).

It is often convenient to refer the thermodynamic properties of a solution to those of an ideal solution. For our present purpose, the simplest definition of an ideal solution is that the heat of mixing is zero, and the entropy of mixing is equal to the configurational entropy of a random arrangement. Consider a simple structure in which  $N$  sites are occupied by  $N_A A$  atoms and  $N_B B$  atoms. The number of distinguishable ways in which the atoms may be arranged on the sites is  $N!/(N_A)!(N_B)!$ , and the configurational entropy of mixing is thus

$$\begin{aligned}\Delta_m S &= k\{\ln N! - \ln N(1-x)! - \ln(Nx)!\} \\ &= -Nk\{(1-x) \ln(1-x) + x \ln x\}\end{aligned}\quad (22.13)$$

and the free energy of mixing per atom is

$$\Delta_m g = kT\{(1-x) \ln(1-x) + x \ln x\}.\quad (22.14)$$

The expression for the entropy of mixing is identical with that obtained previously for lattice defects, which were assumed to be randomly distributed.

In eqns (22.13) and (22.14),  $x$  and  $(1-x)$  are necessarily fractional quantities, less than unity. The free energy of mixing,  $\Delta_m g$ , is thus always negative; it is plotted in Fig. 6.5. The slope of the curve is infinite at  $x = 0$  and  $x = 1$ , and this is sometimes used to justify the statement that pure components cannot form equilibrium phases in an assembly containing two or more components. Very small amounts of a second element dissolved in the lattice will always be distributed randomly, and so give an entropy of mixing corresponding to that of an ideal solution. In a finite crystal, however, solution must be discontinuous, i.e. atom by atom, and there is no reason in principle why the addition of only one atom should not raise the internal energy sufficiently to overcome the associated rise in entropy. This argument is rather academic, and in practice the mixing energy will almost certainly always lead to the presence of a small number of solute atoms in solution at equilibrium, just as

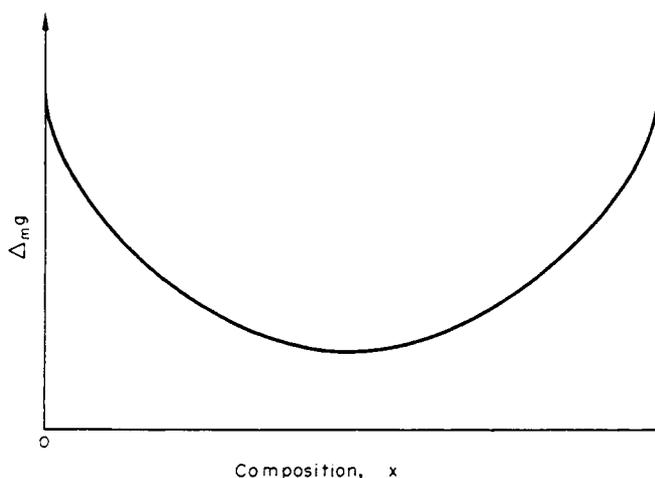


FIG. 6.5. Free energy of mixing vs. composition curve for an ideal solution.

there is always a small number of lattice vacancies. For practical purposes, the solubility may be so small as to be negligible, and we are justified in expecting some solubilities to be too small to be measured by the most sensitive techniques available.

From eqns. (22.6) and (22.4) we may now write expressions for the chemical potentials per atom of an ideal solution. Leaving out the phase identifying superscript,  $\alpha$  or  $\beta$ , since this is common to all quantities, we have

$$\left. \begin{aligned} g_A - g_A^0 &= kT \ln(1-x), \\ g_B - g_B^0 &= kT \ln x. \end{aligned} \right\} \quad (22.15)$$

Guggenheim, and other workers, use a function called the absolute activity,  $\lambda$ , which is related to the chemical potential by the equation  $g_A = kT \ln \lambda_A$ . In terms of these absolute activities, eqn. (22.15) becomes

$$\lambda_A/\lambda_A^0 = 1-x, \quad \lambda_B/\lambda_B^0 = x. \quad (22.16)$$

This equation is useful because it forms the basis of one convenient way of defining the thermodynamic properties of a real solution by means of activity coefficients. For a non-ideal solution, we write

$$\lambda_A/\lambda_A^0 = \gamma_A(1-x), \quad \lambda_B/\lambda_B^0 = \gamma_B x. \quad (22.17)$$

The activity coefficients  $\gamma_A$  and  $\gamma_B$  are functions of composition, and may be determined experimentally for a real solution.

An alternative way of measuring the deviations from "ideal" behaviour is in terms of the "excess" thermodynamic functions. For example, the free energy of mixing may be written

$$\Delta_m g = \Delta_e g + kT\{(1-x) \ln(1-x) + x \ln x\}, \quad (22.18)$$

where  $\Delta_e g$  is called the excess free energy of mixing. The equation

$$\Delta_e g = \Delta_m h - T \Delta_e s \quad (22.19)$$

then defines the (excess) heat of mixing, and the excess entropy of mixing.

The excess quantities just introduced are the mean values per atom. The corresponding partial functions may also be defined in the usual way. For example

$$\Delta_m h = (1-x) \Delta_m h_A + x \Delta_m h_B, \quad (22.20)$$

where  $\Delta_m h_A$ ,  $\Delta_m h_B$  are the partial heats of mixing per  $A$  atom and per  $B$  atom respectively. Thus  $\Delta_m h_A = \partial(N \Delta_m h) / \partial N_A$  is the heat absorbed per atom of  $A$  when a small quantity of  $A$  is added to a very large quantity of solution at constant temperature and pressure; for this reason, it is often called the heat of solution of component  $A$ . Similarly, the excess entropy may be written

$$\Delta_e s = (1-x) \Delta_e s_A + x \Delta_e s_B, \quad (22.21)$$

where  $\Delta_e s_A$ ,  $\Delta_e s_B$  are the excess partial entropies of mixing, or entropies of solution.

In a very dilute solution, the partial excess quantities of the solvent must tend to zero. Thus, there is no heat of solution when  $A$  in the  $\alpha$  phase is dissolved in an  $\alpha$  solution which is already nearly entirely  $A$ , and the entropy of mixing will be represented completely by the configurational term in (22.18). It follows that for a dilute solution

$$\Delta_e g = x(\Delta_m h_B - T \Delta_e s_B) \quad (x \ll 1). \quad (22.22)$$

The  $B$  atoms are sufficiently separated to be considered as non-interacting, and each has associated with it an intrinsic energy, which is the limiting value of  $\Delta_m h_B$  in very dilute solution, and an intrinsic entropy, which is the corresponding limiting value of  $\Delta_e s_B$ . Both of these terms are due essentially to the localized disturbance round each solute atom, and the excess entropy of mixing is entirely due to thermal entropy. The situation in a very dilute solution is exactly analogous to that of a metal containing lattice defects, and the term  $\Delta_m h_B$  corresponds to the intrinsic energy of a defect (e.g.  $\Delta h_{\square}$ ), and  $\Delta_e s_B$  corresponds to the intrinsic thermal entropy of a defect (e.g.  $\Delta s_{\square}$ ).

For very dilute solutions in which  $x_1 \ll 1$ ,  $(1-x_2) \ll 1$ , the equilibrium condition (eqns. (22.4) or (22.5)) becomes

$$(\partial \Delta_m g / \partial x)_{x_1} = (\partial \Delta_m g / \partial x)_{x_2} = 0,$$

so that the variation of the solubility of  $B$  in  $A$  when this solubility is very small is obtained from (22.18) as

$$x_1 = \exp(\Delta_e s_B / k) \exp(-\Delta_m h_B / kT) \quad (22.23)$$

with a similar expression for  $x_2$ .

A similar treatment may be given for the solubility limits in a simple eutectic assembly in which the component metals have different crystal structures, and are very sparingly soluble in each other in the solid. We have now to use the free energies of mixing defined

relative to  $A$  in the  $\alpha$  phase and  $B$  in the  $\beta$  phase (eqn. (22.11)), and the corresponding excess functions which are defined by

$$\Delta_n g = \Delta_f g + kT\{(1-x) \ln(1-x) + x \ln x\}. \quad (22.24)$$

The quantities  $\Delta_n g^\alpha$ ,  $\Delta_f g^\alpha$  differ from  $\Delta_m g^\alpha$  and  $\Delta_e g^\alpha$  respectively by  $x(g_B^{\alpha 0} - g_B^{\beta 0}) = x \Delta g_B^{\beta\alpha}$ , where  $\Delta g_B^{\beta\alpha}$  is the free energy per atom required to change pure  $B$  from the  $\beta$  phase to the  $\alpha$  phase. Similarly,  $\Delta_n g^\beta$  and  $\Delta_f g^\beta$  differ from  $\Delta_m g^\beta$  and  $\Delta_e g^\beta$  respectively by  $(1-x) \Delta g_A^{\alpha\beta}$ .

Writing

$$\Delta_f g^\alpha = x(\Delta_n h_B - T \Delta_f s_B)$$

and using eqn. (22.12), it is readily seen that in very dilute solution the solubility becomes

$$x^\alpha = \exp(\Delta_f s_B/k) \exp(-\Delta_n h_B/kT) \quad (x^\alpha \ll 1), \quad (22.25)$$

which is of the same form as (22.23), and differs only in that

$$\Delta_n h_B = \Delta_m h_B + \Delta h_B^{\beta\alpha}, \quad \Delta_f s_B = \Delta_e s_B + \Delta s_B^{\beta\alpha}. \quad (22.26)$$

It is also possible to obtain an equation of the form (22.23) for the solubility limit when a primary solid solution is in equilibrium with an intermetallic compound; the same procedure is followed, but the free energy of mixing and the corresponding excess functions are defined with reference to the pure metal and the compound as standard state.

Equations (22.23) or (22.25) are only valid for very dilute solutions, when each solute atom has an intrinsic energy and entropy. For moderately dilute solutions ( $x \gtrsim 0.01$ ) Freedman and Nowick (1958) have developed a more general expression by treating eqn. (22.22) as the first term of a Taylor expansion for both  $\Delta_m h$  and  $\Delta_e s$ . For the case of terminal  $\alpha$  and  $\beta$  solutions, they obtain the result

$$\frac{1}{(1-2x^\alpha)} \ln \left\{ \frac{x^\alpha}{1-x^\beta} \right\} = \frac{\Delta_f s_B}{k} - \frac{\Delta_n h_B}{kT}, \quad (22.27)$$

and the equivalent result for two solid solutions of the same structure is clearly obtained by writing  $x^\alpha = x_1$ ,  $x^\beta = x_2$ ,  $\Delta_n h_B = \Delta_m h_B$  and  $\Delta_f s_B = \Delta_e s_B$ . In using this expression to investigate experimental solubility data, the left-hand side is equated to  $\ln x_{\text{corr}}$  thus reducing it to the form of (22.23). A plot of  $\ln x_{\text{corr}}$  against  $1/T$  then enables  $\Delta_n h_B$  and  $\Delta_f s_B$  to be deduced from the slope and the intercept on the  $1/T = 0$  axis of the straight line which should be obtained. In practice, the factor  $1-x^\beta$  may be equated to unity without affecting the results, but the  $1-2x^\alpha$  correction cannot be ignored even in quite dilute solutions, and much better straight lines are obtained from  $\ln x_{\text{corr}}$  than from  $\ln x$ .

An important effect in theories of nucleation (Chapter 10) and growth (Chapter 11) arises from the presence of interfaces in a two-phase assembly. The solubility limits calculated above are strictly valid only for planar interfaces; when interfaces are curved there are additional energy terms which can not be neglected. Consider an assembly in which  $\beta$  particles of surface area  $O$  and surface free energy per unit area  $\sigma$  are in equilibrium with the  $\alpha$  matrix. The energy of the interfaces displaces the equilibrium condition (22.8) or (22.12) and may be considered to contribute an additional thermodynamic potential so

that both the solubility limit  $x^\alpha$  and the equilibrium composition  $x^\beta$  of the  $\beta$  phase vary with the size of the particles. We now denote these quantities as  $x_r^\alpha$  and  $x_r^\beta$  and the equilibrium compositions for infinite planar interfaces by  $x_\infty^\alpha$  and  $x_\infty^\beta$ . Most published treatments of this size effect (Gibbs–Thomson effect) are oversimplified and assume that  $x^\beta$  remains constant.

In a virtual change in which  $dn$  atoms are transferred from the  $\alpha$  phase to the  $\beta$  phase, there is now an additional energy term  $\sigma dO$  due to the increase in surface area of the  $\beta$  particles. As shown in Fig. 6.6, this may be represented as a displacement of  $g^\beta$  to  $g_r^\beta = g^\beta + \sigma(dO/dn)$  and the new equilibrium compositions are given by the points of contact of the tangent common to the curves of  $g^\alpha$  and  $g_r^\beta$  against  $x$ . Let the effective chemical potentials per atom for curved interfaces, defined by the construction of Fig. 6.2 applied

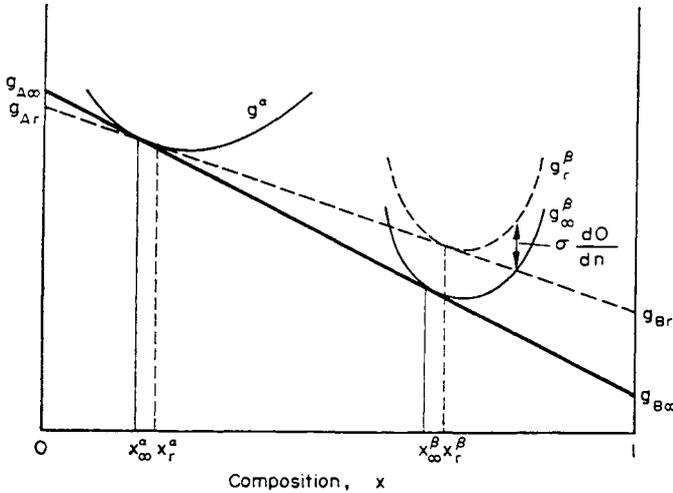


FIG. 6.6. To illustrate the Gibbs–Thomson effect.

to  $g_r^\beta$ , be denoted  $g_{Ar}^\beta$ ,  $g_{Br}^\beta$ . The new equilibrium conditions are  $g_A^\alpha(x_r^\alpha) = g_{Ar}^\beta(x_r^\beta)$  and  $g_B^\alpha(x_r^\alpha) = g_{Br}^\beta(x_r^\beta)$  (compare eqn. (22.8)), and we write this more briefly as  $g_{Ar}^\alpha = g_{Ar}^\beta$ , etc. From the geometry of Fig. 6.6, we note that

$$\Delta x_r^\alpha = x_r^\alpha - x_\infty^\alpha \tag{22.28}$$

has the same sign as  $\Delta x_r^\beta$  and is of comparable magnitude, and also that to a good approximation

$$g_{Br}^\alpha - g_{B\infty}^\alpha = \{(1 - x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha)\} \sigma(dO/dn) \tag{22.29}$$

In terms of the activity coefficients (eqn. (22.17)), we also have

$$g_{Br}^\alpha - g_{B\infty}^\alpha = kT \ln(\gamma_{Br}^\alpha x_r^\alpha / \gamma_{B\infty}^\alpha x_\infty^\alpha) \tag{22.30}$$

and for dilute solutions or solutions which obey Henry's law  $\gamma_B^\alpha$  is a constant. Equations (22.29) and (22.30) thus give

$$\frac{\Delta x_r^\alpha}{x_\infty^\alpha} \approx \ln \frac{x_r^\alpha}{x_\infty^\alpha} = \frac{\sigma}{kT} \left( \frac{dO}{dn} \right) \left( \frac{1 - x_\infty^\alpha}{x_\infty^\beta - x_\infty^\alpha} \right). \tag{22.31}$$

In the often used approximation of spherical particles

$$dO/dn = 2v^\beta/r \quad (22.32)$$

and

$$\Delta x_r^\alpha = (2\Gamma/r)x_\infty^\alpha, \quad (22.33)$$

where

$$\Gamma = (\sigma v^\beta/kT)(1-x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha). \quad (22.34)$$

For sufficiently small values of  $x_\infty^\alpha$  (dilute solutions)

$$\Gamma = (\sigma v^\beta/kT)/(x_\infty^\beta - x_\infty^\alpha), \quad (22.35)$$

and this form which has been used by Trivedi (1970) in theories of diffusion-controlled growth (see Chapter 11). A still more extreme assumption is that  $x^\beta = 1$  in which case

$$\Gamma = \sigma v^\beta/kT. \quad (22.36)$$

Equations (22.33) and (22.36) together constitute a frequently quoted (but obviously inaccurate) statement of the Gibbs–Thomson theory. Finally, in the general case where Henry's law does not apply,  $\Gamma$  in eqn. (22.33) is given by

$$\Gamma = (\sigma v^\beta/kT)(1-x_\infty^\alpha)/(x_\infty^\beta - x_\infty^\alpha) \{1 + (d \ln \gamma_B^\alpha / d \ln x)\}. \quad (22.37)$$

Many results obtained with the special assumption of eqn. (22.36) remain valid under more general conditions if  $\Gamma$  from (22.37) is used in place of  $(\sigma v^\beta/kT)$  in the appropriate equations.

If the particles are cylinders (needles) rather than spheres  $dO/dn = v^\beta/r$  and eqn. (22.33) is modified to

$$\Delta x_r^\alpha = (\Gamma/r)x_\infty^\alpha. \quad (22.38)$$

In the general case of equilibrium at a curved interface

$$dO/dn = v^\beta(r_1^{-1} + r_2^{-1}), \quad (22.39)$$

where  $r_1$  and  $r_2$  are the principal radii of curvature. Similarly, for polyhedral particles of equilibrium shape, the surface energy term becomes  $d(O_i\sigma_i)/dn$  where  $O_i$  is a surface in the equilibrium shape at distance  $h_i$  from the centre of the particle so that according to the theory of Section 20,  $\sigma_i/h_i$  is constant. This gives a more general form of (22.33) as

$$\Delta x_r^\alpha = (2\sigma_i/h_i)(\Gamma/\sigma)x_\infty^\alpha. \quad (22.40)$$

Note that since  $\sigma$  is positive  $x_r^\alpha$  is always greater than  $x_\infty^\alpha$ . We may usefully digress here to point out that there is an equivalent relationship between the vapour pressure in equilibrium with a liquid droplet of radius  $r$  and that in equilibrium with a flat liquid surface (Chapter 10, eqn. (46.14)). The curvature of a surface may in fact be regarded as introducing a pressure differential

$$p = \sigma(r_1^{-1} + r_2^{-1}) \quad (22.41)$$

since for a fluid medium the surface tension is numerically equal to the surface free energy  $\sigma$  (Section 35). This concept leads to the same expression for the change in solubility with particle size since the free energy change per atom is now  $pv^\beta$ .

Equations (22.31), (22.33), (22.37), and (22.40) and the equivalent vapour pressure eqn. (46.14) are variously known as the Gibbs–Thomson, Thomson–Helmholtz, or Thomson–Freundlich equation (Gibbs 1878; Thomson 1888; Freundlich 1926) or sometimes simply as the effects of “capillarity”, which is the term used by Gibbs.<sup>†</sup> The equations are often written in terms of concentrations  $c_{\infty}^{\alpha}$ ,  $c_r^{\alpha}$ , etc., instead of atomic fractions  $x_{\infty}^{\alpha}$ ,  $x_r^{\alpha}$ , etc, but this is strictly valid only when there is negligible volume change on transformation. The above treatment is implicitly based on this assumption since no allowance is made for strain energies due to transformation (see below).

We conclude this section by examining the thermodynamics of inhomogeneous phases and the criteria for distinguishing between a metastable phase and an unstable phase. A solid solution of fluctuating composition may be assigned a free energy which contains three different terms. The first contribution is the sum of the free energies which the individual atoms would have if they were present in homogeneous solutions of the same compositions as those of their local environments. Each term in the sum is thus represented by a point on the free-energy curve of the homogeneous solution. In the particular case that there are regions of only two different compositions, the change in free energy per atom in comparison with the homogeneous solution is given graphically by the construction shown in Fig. 6.1, and the sum over  $N$  atoms is just a generalization of this procedure. The second term arises because the free energy of any small volume element of given composition will be changed if it is surrounded by material of different composition. When the transition is sharp, the extra energy appears macroscopically as the surface free energy we have discussed above; Cahn and Hilliard (1958) showed how to express the analogous energy which arises from a continually fluctuating composition. The third term was mentioned briefly above; it is the elastic coherency strain energy which is due to a variation of lattice parameter with composition. If all regions of the inhomogeneous phase are coherent with each other, the assembly will be in a self-stressed state, the elastic energy of which must be considered part of the overall free energy.

We consider first the chemical or volume free energy. For *small* variations about the mean composition, the generalized construction of Fig. 6.1 will always leave  $R$  above  $P$  in those composition ranges where  $(\partial^2 g/\partial x^2) > 0$ , and this includes some compositions between  $x_1$  and  $x_2$ . However, there is a net decrease in free energy, even for very small fluctuations, if  $(\partial^2 g/\partial x^2) < 0$ . The curve

$$(\partial^2 g/\partial x^2) = (\partial^2 \Delta_m g/\partial x^2) = 0 \quad (22.42)$$

was called by Gibbs the limit of metastability, since inside this curve a homogeneous solution should decompose spontaneously at a rate limited only by the rate of atomic migration. The curve is now usually referred to as the spinodal, or more strictly as the chemical spinodal. As we shall describe in later chapters, the decomposition of a supersaturated solid solution inside the spinodal does not require nucleation in the classical sense and occurs by a process known as “uphill” or “negative” diffusion. Note that there is no spinodal in free-energy diagrams such as those shown in Figs. 6.2–6.5.

<sup>†</sup> An interesting survey of the early history of this subject is given by Swan and Urquhart (1927).

Cahn and Hilliard used a multivariable Taylor expansion to express the free energy of an inhomogeneous solution in terms of  $g^a$  and of the various spatial derivatives of composition  $x$  (or of concentration  $c$ ). The first term in this expansion is the sum already considered and for centrosymmetric crystals the energy per atom in the next most important term may be written as  $Kv(\nabla x)^2 = Kv^3(\nabla c)^2$ , where  $v$  is the atomic volume and  $K$  is a second rank tensor and is thus isotropic for cubic crystals. This "gradient energy" is a continuum analogue of the surface free energy of a discontinuous interface; the Cahn and Hilliard expansion is, of course, invalid if the gradient  $\nabla c$  becomes large. There has been some dispute over the validity of the symmetry argument for the absence of a term in  $|\nabla c|$  (see Tiller *et al.* 1970; Cahn and Hilliard, 1971), but we believe the Cahn and Hilliard treatment is correct.

An exact theory of the coherency energy is difficult when allowance is made for elastic anisotropy (see Cahn 1969), and we mention here only the isotropic result. The elastic energy depends on  $\varepsilon = d \ln a/dx$ , where  $a$  is the lattice parameter, and on an elastic modulus  $Y' = Y/(1-\nu)$ , where  $Y$  is Young's modulus and  $\nu$  is Poisson's ratio. Cahn (1961) showed that if the actual composition is  $x$  and the mean composition is  $x_0$  the coherency energy per atom may be written as  $\varepsilon^2 Y' v (x - x_0)^2$ . This changes both the solubility limits and the limits of metastability. In place of eqn. (22.4), the coherent miscibility gap is defined by  $x_{1c}$  and  $x_{2c}$ , where

$$\left( \frac{\partial g^a}{\partial x} \right)_{x_{1c}} = \frac{g(x_{2c}) - g(x_{1c})}{x_{2c} - x_{1c}} + 2v\varepsilon^2 Y' (x_{2c} - x_{1c}) \quad (22.43)$$

with a similar equation for  $x_{2c}$  and

$$x_{1c} > x_1, \quad x_{2c} < x_2. \quad (22.44)$$

Similarly, the coherent spinodal which represents the true limit of metastability is defined by

$$(\partial^2 g / \partial x^2) + 2\varepsilon^2 Y' v = 0, \quad (22.45)$$

and lies entirely inside the chemical spinodal. The increase in solubility  $x_{1c} - x_1$  at fixed temperature, or equivalently the depression of the temperature of the coherent curves below the chemical curve, depends sensitively on the magnitude of  $\varepsilon$ ; the supercooling from the chemical to the coherent spinodal was estimated by Cahn as 40°C for aluminium-zinc alloys and 200°C for gold-nickel alloys.

Note that although the coherency strain energy is a function of the mean composition  $x_0$  of the alloy, the positions of the coherent phase boundary and spinodal are not dependent on  $x_0$ . However, directional effects appear when anisotropic elastic theory is used; for example, the solid solution may be unstable to fluctuations in composition in some directions in a crystal but metastable for other directions (see Part II, Chapter 18).

Equation (22.44) shows that the coherent spinodal only exists when there is a negative region of  $(\partial^2 g / \partial x^2)$ , i.e. when there is also a chemical spinodal. However, precipitates of one phase in another which involve chemical free energy curves of the type shown in Fig. 6.3 may sometimes be obtained in either coherent or incoherent forms, and there is then in principle both a coherent solvus and an equilibrium solvus. Intermediate or metastable phases which are coherent with the matrix are also often encountered in low-temperature

precipitation reactions. Such phases frequently have a lower symmetry than the equilibrium phase and it may occasionally be illuminating to regard them as coherent versions of the equilibrium precipitate.

### 23. THE NEAREST NEIGHBOUR MODEL : REGULAR SOLUTIONS

The forces between the atoms in a solid solution, as in a pure metal, are mainly short-range in character, and the most used model is that in which only nearest neighbour interactions are considered significant. The energy of the crystal is then the sum of the pair interaction energies of the  $A-A$ ,  $B-B$ , and  $A-B$  contacts. This is often referred to as the Ising model, although Ising actually treated only the magnetic analogue of this situation; the first application to alloys was made by Bethe.

The limitations of the nearest neighbour model in providing a satisfactory description of the metallic state have already been emphasized in the previous chapter. When applied to solid solutions, an additional difficulty is at once apparent, since the theory predicts that all properties of the solid solution are symmetrical about the 50 atomic% composition, and this is rarely true in practice. This limitation may be avoided by modifying the theory, but at least two parameters are then required to specify the energy of the assembly. A major attraction of the simple central force model (to the mathematician, at least!) is that all properties of the solid solution at fixed temperature, and of the equilibrium diagram, can be described in terms of a single arbitrary parameter.

Let us consider a binary alloy containing  $A$  and  $B$  atoms. In place of the characteristic nearest neighbour interaction energy,  $-2\bar{\mathcal{E}}$ , we now have the energies  $-2\mathcal{E}_{AA}$ ,  $-2\mathcal{E}_{BB}$ , and  $-2\mathcal{E}_{AB}$  representing the binding energies of two  $A$  atoms, two  $B$  atoms and an  $A$  atom and a  $B$  atom respectively. The model assumes that  $\mathcal{E}_{AA}$ ,  $\mathcal{E}_{BB}$ , and  $\mathcal{E}_{AB}$  are all independent of the surrounding configuration, and this is probably its most serious limitation.

The three potential energy terms are independent of each other, but in the nearest neighbour model, the properties of the solid solution depend only on the combination

$$\bar{\mathcal{E}} = \mathcal{E}_{AA} + \mathcal{E}_{BB} - 2\mathcal{E}_{AB}, \quad (23.1)$$

The nature of the quantity  $\bar{\mathcal{E}}$  may be seen by a hypothetical process in which we interchange an  $A$  atom and a  $B$  atom on any two sites in the crystal. If the number of  $A-A$  contacts increases or decreases by  $X$ , it follows that whatever the arrangement around the two sites, the number of  $B-B$  contacts also changes by  $X$ , and the number of  $A-B$  contacts by  $-2X$ . The change in energy is thus  $-X(\mathcal{E}_{AA} + \mathcal{E}_{BB} - 2\mathcal{E}_{AB})$ .

In an ideal solid solution, the internal energy of the crystal is independent of the atomic arrangement. This condition is satisfied if  $\mathcal{E}_{AA} = \mathcal{E}_{BB} = \mathcal{E}_{AB}$ , and (less restrictively) if  $\bar{\mathcal{E}} = 0$ . The condition for the formation of an ideal solid solution with this model is thus that the force between unlike atoms is equal to the average of the forces between two like atoms. When longer-range forces are considered, the corresponding conditions are obvious; in any interchange of  $A$  and  $B$  atoms within the interior of the crystal, the change in energy must vanish. One way of including longer-range forces is to use a pairwise interaction model

with a series of interchange potentials  $\mathcal{E}_1, \mathcal{E}_2$ , etc., defined by

$$\mathcal{E}_i = \mathcal{E}_{AA,i} + \mathcal{E}_{BB,i} - 2\mathcal{E}_{AB,i} \quad (23.1a)$$

where each potential is defined at a separation corresponding to the  $i$ th nearest neighbour distance. Note that although, according to this model, there are three interaction potentials for each distance of separation, it is impossible to imagine an experiment which would enable the individual potential to be measured separately, and only the interchange potentials  $\mathcal{E}_i$  have operational significance. In any interchange of an  $A$  atom and a  $B$  atom, the change in energy is now given by the sum  $-X_i\mathcal{E}_i$ , where  $X_i$  is the increase in the number of  $i$ th neighbour  $A$ - $A$  or  $B$ - $B$  bonds, and  $-2X_i$  is the corresponding increase in  $A$ - $B$  bonds.

For many solid solutions of close-packed structures, the limitations of the nearest neighbour model lie more in the assumption of central forces and the independence of the  $\mathcal{E}$  terms on the environment than in the neglect of second and third nearest neighbour interactions. Thus Guggenheim (1952) showed that if the interaction energy varies as  $r^{-6}$ , the effect of second nearest neighbours on measurable thermodynamic quantities is negligible. The use of such interactions makes equations unwieldy and only in special circumstances does it lead to any essential improvement in the physical description. We shall therefore confine ourselves to nearest neighbour forces in this section, although there are then obvious difficulties in applying the theory to b.c.c. structures which are not mechanically stable under nearest neighbour forces. It will be necessary to consider higher neighbour interactions in connection with order-disorder changes (Section 26), and some comments will also be made there on their influence in systems exhibiting phase segregation.

Non-ideal solutions, which have  $\mathcal{E} \neq 0$ , may be classified qualitatively by the sign of  $\mathcal{E}$ . If  $\mathcal{E}_{AA} + \mathcal{E}_{BB} < 2\mathcal{E}_{AB}$ , the attractive forces between like atoms are weaker than those between unlike atoms, and there will be a tendency for each atom to surround itself with as many atoms of the opposite kind as possible. This ordering tendency may produce a superlattice at low temperatures; at high temperatures, it is opposed by the thermal energy, which always tends to produce a random arrangement of high entropy. For  $\mathcal{E}$  positive, the opposite result is valid, and at low temperatures the solid solution tends to segregate into  $A$ -rich and  $B$ -rich regions.

When the solution is not ideal, theoretical expressions can most readily be obtained on the assumption that it is "regular". This term was first introduced by Hildebrand to describe a class of solutions having physical properties which vary with composition in a regular manner; the definition of the regular solution has, however, varied considerably amongst different workers. The earliest approach was to define a regular solution as a solution in which the configurational entropy of mixing is still given by eqn. (22.13), even though the heat of mixing is not zero. The atomic arrangement is thus considered to be effectively random, although interchanges of atoms lead to changes in the internal energy. This assumption is obviously roughly justified if the magnitude of  $\mathcal{E}$  is small. In this sense, a regular solution is one which deviates only slightly from ideal conditions.

Guggenheim (1952) uses a less restricted definition of regular solution, and regards the above definition as a crude or "zeroth order" approximation. This approach has some ad-

vantage in emphasizing the relation between the simple theory and the higher approximations of the quasi-chemical theory.

For a simple structure in which each site has a common coordination number  $z$ , there will be  $\frac{1}{2}Nz$  nearest neighbour bonds in a crystal of  $N$  atoms, provided  $N$  is sufficiently large for surface effects to be negligible. We write the total number of nearest neighbour  $A-B$  contacts as  $zN_{AB}$  (i.e.  $zN_{AB}$  is the value of  $N_{AB}(\mathbf{r})$  for the whole crystal when  $\mathbf{r}$  is a nearest neighbour vector).<sup>†</sup> Similarly,  $zN_{AA}$  and  $zN_{BB}$  are the total numbers of  $A-A$  and  $B-B$  contacts in the crystal. By considering the neighbours of the  $A$  atoms and the  $B$  atoms separately, we find

$$\left. \begin{aligned} \text{Number of } A-A \text{ contacts} &= zN_{AA} = \frac{1}{2}z(N_A - N_{AB}), \\ \text{Number of } B-B \text{ contacts} &= zN_{BB} = \frac{1}{2}z(N_B - N_{AB}), \\ \text{Number of } A-B \text{ contacts} &= zN_{AB}. \end{aligned} \right\} \quad (23.2)$$

For a random distribution,  $P_{AA}(\mathbf{r}) = (1-x)^2$ ,  $P_{BB}(\mathbf{r}) = x^2$ , and  $P_{AB}(\mathbf{r}) = 2x(1-x)$  for all possible interatomic vectors  $\mathbf{r}$ . In particular, the numbers of nearest neighbour pairs of the types  $A-A$ ,  $B-B$ ,  $A-B$  will be  $\frac{1}{2}Nz(1-x)^2$ ,  $\frac{1}{2}Nzx^2$ , and  $Nzx(1-x)$ . The potential energy of the crystal is thus

$$\left. \begin{aligned} U &= -Nz\{(1-x)^2\mathcal{E}_{AA} + x^2\mathcal{E}_{BB} + 2x(1-x)\mathcal{E}_{AB}\} \\ &= -Nz\{(1-x)\mathcal{E}_{AA} + x\mathcal{E}_{BB} - x(1-x)\mathcal{E}\}. \end{aligned} \right\} \quad (23.3)$$

The value of  $U$  for the components is  $-Nz(1-x)\mathcal{E}_{AA} - Nz x\mathcal{E}_{BB}$ , so that we have

$$\Delta_m H = \Delta_m U = Nz x(1-x)\mathcal{E} \quad (23.4)$$

as the heat or energy of mixing.<sup>‡</sup> This curve has a simple  $U$ -shape, with a maximum or minimum at  $x = \frac{1}{2}$ , according to the sign of  $\mathcal{E}$ . The configurational entropy of mixing is given by eqn. (22.13), since the atomic arrangement is assumed to be random. If we assume for simplicity that the thermal entropy of mixing is zero, we obtain for the free energy of mixing per atom

$$\Delta_m g = zx(1-x)\mathcal{E} + kT\{(1-x)\ln(1-x) + x\ln x\}. \quad (23.5)$$

The form of the curve of  $\Delta_m g$  against composition thus depends on the sign of  $\mathcal{E}$ . The relation is shown in Fig. 6.7 for  $\mathcal{E}$  negative (tendency to form superlattices),  $\mathcal{E}$  zero (ideal solutions) and  $\mathcal{E}$  positive (tendency for phase segregation). The latter case corresponds to Fig. 6.1, and the compositions of the two phases in equilibrium are given by eqn. (22.5). In the present model, eqn. (23.5) is symmetrical about  $x = \frac{1}{2}$ , and the common tangent to the two minima in the curve has zero slope, so that

$$(\partial\Delta_m g/\partial x)_{x_1} = (\partial\Delta_m g/\partial x)_{x_2} = 0. \quad (23.6)$$

<sup>†</sup> It is more convenient to write the number of  $A-B$  contacts as  $zN_{AB}$  than as  $N_{AB}$ ; this means that  $P_{AB}(\mathbf{r}) = 2N_{AB}/N$  for nearest neighbour contacts, and simplifies the algebra of the quasi-chemical theory.

<sup>‡</sup> In condensed phases, we need not distinguish between  $\Delta U$  and  $\Delta H$ , or between  $\Delta F$  and  $\Delta G$ .

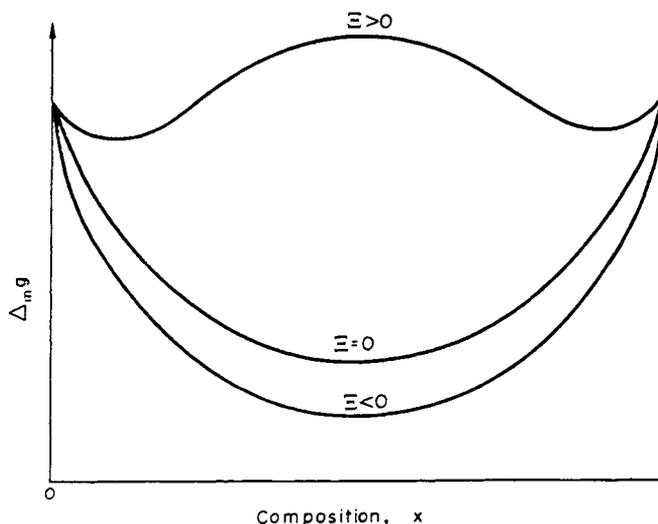


FIG. 6.7. Free energy of mixing vs. composition curves according to (23.5).

From eqns. (23.5) and (23.6),

$$\left(\frac{1}{2x-1}\right) \ln\left(\frac{x}{1-x}\right) = \frac{z\Xi}{kT}. \quad (23.7)$$

At temperatures where the curve of  $\Delta_m g$  has the form of Fig. 6.1, there are three solutions to (23.7). The solution at  $x = \frac{1}{2}$  gives a maximum value to  $\Delta_m g$ ; the other two solutions,  $x_1$  and  $x_2$ , give minima, and are symmetrical about  $x = \frac{1}{2}$ . For dilute solutions ( $x \ll 1$ ), we write as an approximation  $(1-x) \approx (1-2x) \approx 1$ , and the variation of the solubility limit  $x_1$  is then

$$x_1 = \exp-(z\Xi/kT), \quad (23.8)$$

so that the solubility approximates to an exponential curve if the solubility is small, i.e. if  $z\Xi/kT$  is large.

The similarity of eqn. (23.7) to Freedman and Nowick's expression eqn. (22.27) should be noted, but the underlying physical assumptions are different. Equation (22.27) applies to any solubility limit if the solubility is small, and includes the effects of thermal entropy. Equation (23.7) applies only to the artificial regular solution model, but within this limitation it is valid for all values of  $x$ . The limiting forms of the two equations in very dilute solutions (eqns. (22.23) and (23.8)) differ insofar as a thermal entropy contribution from each solute atom is present in a real solution, but not in the model.

Equation (23.7) is sometimes written in the form

$$(2/(2x-1)) \tanh^{-1}(2x-1) = z\Xi/kT. \quad (23.9)$$

As the temperature is raised, the two roots  $x_1$  and  $x_2$  both approach the value  $x = \frac{1}{2}$  until, finally, they coincide at this value, and the curve of  $\Delta_m g$  assumes a simple U-form. At all

temperatures above this critical temperature, there is a continuous solid solution from pure  $A$  to pure  $B$ . The  $\Delta_m g$  curve has a maximum at  $x = \frac{1}{2}$  for  $T < T_c$ , and a minimum at  $x = \frac{1}{2}$  for  $T > T_c$ , so that  $T_c$  is given by the condition  $(\partial^2 \Delta_m g / \partial x^2)_{x=1/2} = 0$ . Using eqn. (23.5),

$$T_c = z\bar{\epsilon}/2k. \quad (23.10)$$

The complete solubility curve for the hypothetical assembly we are considering is plotted in Fig. 6.8, the temperatures being measured in units of  $kT/z\bar{\epsilon}$ . The assumptions lead to an equilibrium diagram in which there is a solubility gap below  $T_c$ , the boundaries of the gap being symmetrical about the equi-atomic composition. The solubilities are zero at 0 K,

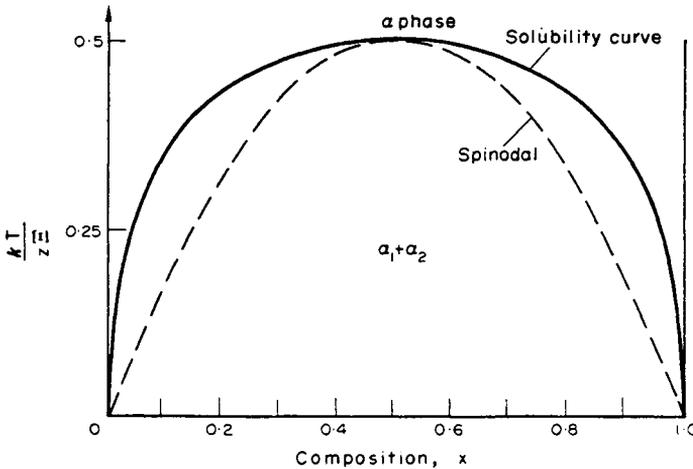


FIG. 6.8. The solubility limits and spinodal curve given by eqns. (23.9) and (23.11).

since the formation of a solid solution raises the internal energy of the assembly above that of a phase mixture of the pure components. As already noted, the equilibrium diagram of Fig. 6.8 is simply related to that of a eutectic assembly in which two metals of the same crystal structure have a limited mutual solubility in the solid state. The eutectic reaction occurs if the liquid phase becomes stable before the solubility gap is closed.

In addition to the boundaries of the two-phase region we are also interested in the chemical spinodal [eqn. (22.42)] which is given by

$$x(1-x) = kT/2z\bar{\epsilon} \quad (23.11)$$

in a regular solution. At all temperatures below  $T_c = z\bar{\epsilon}/2k$ , the roots of this equation lie inside the roots  $x_1$  and  $x_2$  of eqn. (23.9). The two curves, which touch at  $T = 0$  K and  $T = T_c$ , are shown in Fig. 6.8.

The chemical potential of  $A$  in the regular solution is given by

$$g_A - g_A^0 = z\bar{\epsilon}x^2 + kT \ln(1-x)$$

from eqn. (23.5). The activity coefficient is thus

$$\gamma_A = \exp(z\bar{\mathcal{E}}x^2/kT), \quad (23.12)$$

and we have, similarly,

$$\gamma_B = \exp\{z\bar{\mathcal{E}}(1-x)^2/kT\}.$$

#### 24. THE QUASI-CHEMICAL THEORY: OTHER STATISTICAL THEORIES

We now consider the possibility of removing some of the very restrictive assumptions used in the zeroth order approximation of the treatment of regular solutions. The simple theory above is incorrect because of the neglect of thermal entropy factors, and in the assumption of random atomic arrangement. Both these factors may be important, but it is difficult to devise a model which will include them together in a satisfactory manner. In the quasi-chemical theory, which we shall describe first, emphasis is placed on improving the calculation of the configurational entropy, and the properties of the solution are again expressed in terms of the single parameter  $\bar{\mathcal{E}}$ . In the usual application,  $\bar{\mathcal{E}}$  is temperature independent, as assumed above, and the only contribution to  $\Delta_m s$  comes from configurational terms. Thermal entropy terms can be formally included if  $\bar{\mathcal{E}}$  is allowed to vary with temperature.

In a real solution, it is physically obvious that the atom distribution will be random only at very high temperatures, and the configurational entropy of mixing must therefore be a function of temperature. Since a random distribution gives the maximum number of distinguishable arrangements which the assembly can possess, a more accurate calculation of the configurational entropy term must lead to a negative excess entropy of mixing if the thermal entropy of mixing is ignored. Consider the assembly in a particular macroscopic configuration which we may specify by the number of nearest neighbour  $A-B$  pairs  $zN_{AB}$ . To find the equilibrium state, we have to calculate the number of microscopic states leading to this configuration, and the energy of the configuration. We can then take the sum over all values of  $N_{AB}$ , and thus obtain the partition function.

The total energy of the assembly is  $U = -z(N_A\bar{\mathcal{E}}_{AA} + N_B\bar{\mathcal{E}}_{BB} - N_{AB}\bar{\mathcal{E}})$  (using eqn. (23.2)), and the heat of mixing is thus

$$\Delta_m H = \Delta_m U = z\overline{N_{AB}}\bar{\mathcal{E}}, \quad (24.1)$$

where  $\overline{N_{AB}}$  is the equilibrium value of  $N_{AB}$ . Denoting the total number of arrangements for given  $N_{AB}$  by  $\Omega(N_A, N_B, N_{AB})$ , we have the partition function

$$Q = \sum_{N_{AB}} \Omega(N_A, N_B, N_{AB}) \exp(-U/kT). \quad (24.2)$$

This partition function has not been evaluated exactly for a three-dimensional lattice. An exact solution in two dimensions was obtained by Onsager (1944) using a complex mathematical method, and his results have been extended by other workers using simpler methods.

In three dimensions, the value of  $\Omega(N_A, N_B, N_{AB})$  may be evaluated if we assume as an approximation that the various kinds of pair do not interfere with each other, and may be

treated as separate entities. We note first of all that

$$\Sigma\Omega(N_A, N_B, N_{AB}) = N!/(N_A)!(N_B)! \quad (24.3)$$

since this is merely the expression for the total number of arrangements of the assembly. We are going to be concerned with the logarithm of the function  $\Omega$ , and we may then replace the sum on the left of this expression by the maximum term; this procedure is very commonly used in statistical thermodynamics. Suppose the maximum value of  $\Omega(N_A, N_B, N_{AB})$  is obtained for some particular value  $N_{AB}^0$ . We then have

$$\Omega(N_A, N_B, N_{AB}^0) = N!/(N_A)!(N_B)!$$

The assumption that the total number of arrangements for given  $N_{AB}$  is obtained by treating the various pairs as independent entities requires that  $\Omega(N_A, N_B, N_{AB})$  is proportional to

$$\frac{(\frac{1}{2}zN)!}{\{\frac{1}{2}z(N_A - N_{AB})\}! \{\frac{1}{2}z(N_B - N_{AB})\}! (\frac{1}{2}zN_{AB})! (\frac{1}{2}zN_{BA})!}$$

in which we have allowed for the orientation of the sites, which effectively distinguishes an  $A-B$  pair from a  $B-A$  pair. The number of arrangements is only proportional to the above expression, not equal to it, since the total number of arrangements must satisfy (24.3). We can achieve this by introducing a normalization factor, so that from (24.3) we find

$$\Omega(N_A, N_B, N_{AB}) = \frac{N!}{N_A! N_B!} \frac{\{\frac{1}{2}z(N_A - N_{AB}^0)\}! \{\frac{1}{2}z(N_B - N_{AB}^0)\}! \{\frac{1}{2}zN_{AB}^0\}! \{\frac{1}{2}zN_{AB}^0\}!}{\{\frac{1}{2}z(N_A - N_{AB})\}! \{\frac{1}{2}z(N_B - N_{AB})\}! \{\frac{1}{2}zN_{AB}\}! \{\frac{1}{2}zN_{AB}\}!} \quad (24.4)$$

Also, by differentiating expression (24.4) with respect to  $N_{AB}$ , we obtain the maximum value of  $N_{AB}$  which is given by

$$0 = \partial\Omega(N_A, N_B, N_{AB})/\partial N_{AB} = \partial \ln \Omega(N_A, N_B, N_{AB})/\partial N_{AB}.$$

On taking logarithms and using Stirling's theorem, this reduces to

$$N_{AB}^0 = N_A N_B / N = Nx(1-x), \quad (24.5)$$

which is the value of  $N_{AB}$  for a completely random arrangement, as anticipated.

We can now rewrite eqn. (24.2) as

$$Q = \Sigma\Omega(N_A, N_B, N_{AB}) \exp\{-z(N_A \bar{\epsilon}_{AA} + N_B \bar{\epsilon}_{BB} + N_{AB} \bar{\epsilon})/kT\}.$$

and we now have an explicit value for  $\Omega$ . Once again we replace the sum by its maximum term; this corresponds to  $N_{AB}$  having its equilibrium value  $\bar{N}_{AB}$ . Thus

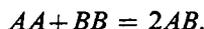
$$Q = \Omega(N_A, N_B, \bar{N}_{AB}) \exp\{-z(N_A \bar{\epsilon}_{AA} + N_B \bar{\epsilon}_{BB} + \bar{N}_{AB} \bar{\epsilon})/kT\} \quad (24.6)$$

and  $\partial \ln Q/\partial N_{AB} = 0$  for  $N_{AB} = \bar{N}_{AB}$ . We have, therefore, to substitute eqn. (24.4) into

(24.6) and differentiate in order to find  $\overline{N_{AB}}$ . Proceeding as we did for  $N_{AB}^0$ , we find that

$$(\overline{N_{AB}})^2 = (N_A - \overline{N_{AB}})(N_B - \overline{N_{AB}}) \exp(-2\Xi/kT). \quad (24.7)$$

This is the basic formula of the present approximation, and was originally derived by Guggenheim by a different method. He treated the interchange of an  $A$  atom and a  $B$  atom as a chemical process in which  $A-A$  bonds and  $B-B$  bonds "react" to form  $A-B$  bonds, and vice versa:



If the law of mass action is applied to this symbolic equation, we obtain the result (24.7). For this reason, the method is often called the quasi-chemical theory. Guggenheim has also shown it is equivalent to the method which Bethe developed for order-disorder reactions (see p. 221).

Equation (24.7) is a rather unpleasant quadratic. The algebra is simplified by writing the solution in the form

$$\overline{N_{AB}} = 2N_{AB}^0/(\beta+1) = 2Nx(1-x)/(\beta+1). \quad (24.8)$$

Substituting this into (24.7) shows that  $\beta$  is the positive root of the equation

$$\beta^2 - (1-2x) = 4x(1-x) \exp(2\Xi/kT). \quad (24.9)$$

Equation (24.8) shows that  $\beta = 1$  in the zeroth approximation, and  $\beta > 1$  in the present treatment. It follows from (24.1) and (24.8) that the heat of mixing per atom is

$$\Delta_m h = \Delta_m u = 2zx(1-x)\Xi/(\beta+1). \quad (24.10)$$

We can now deduce the other thermodynamic properties of the assembly, using the relation

$$G = F = -kT \ln Q = -kT \ln \Omega(N_A, N_B, \overline{N_{AB}}) + U.$$

From eqn. (24.6) it follows that  $\partial F/\partial \overline{N_{AB}} = 0$ , and we already know that  $\partial \ln \Omega/\partial N_{AB}^0 = 0$ .

We may also write:

$$\begin{aligned} \Delta_m g &= x_A(g_A - g_A^0) + x_B(g_B - g_B^0) = kT \left[ (1-x) \ln(1-x) + x \ln x \right. \\ &\quad \left. + \frac{1}{2} z \left\{ x_A \ln \left( \frac{N_A - \overline{N_{AB}}}{N_A - N_{AB}^0} \right) + x_B \ln \left( \frac{N_B - \overline{N_{AB}}}{N_B - N_{AB}^0} \right) \right\} \right] \\ &\quad + 2zx(1-x)\Xi/(\beta+1). \end{aligned} \quad (24.11)$$

This equation is simple and symmetrical.† The quantities  $\overline{N_{AB}}$  and  $N_{AB}^0$  are defined in eqn. (24.7) and (24.5); as already noted, they have simple physical interpretations.  $N_{AB}^0$  is the number of  $A-B$  bonds for a random mixture, and  $\overline{N_{AB}}$  is the mean number of  $A-B$  bonds

† The internal energy  $U$  was inadvertently left out of eqn. 24.11 (24.13 in previous printings). I am grateful to Professor H.K.D.H. Bhadeshia for pointing this out to me.

for the actual solution. If we wish to eliminate the numbers  $N_A$ ,  $N_B$ ,  $\overline{N_{AB}}$ , we may write the equation in the form

$$\begin{aligned} \Delta_m g = kT & \left[ (1-x) \ln(1-x) + x \ln x \right. \\ & \left. + \frac{1}{2} z \left\{ (1-x) \ln \frac{\beta+1-2x}{(1-x)(\beta+1)} + x \ln \frac{\beta-1+2x}{x(\beta+1)} \right\} \right] \\ & + 2zx(1-x)\mathcal{E}/(\beta+1). \end{aligned} \quad (24.12)$$

From eqns. (24.10) and (24.12) an expression for the excess entropy of mixing may be obtained.

To find the equilibrium state, we must equate  $\partial\Delta_m g/\partial x = 0$ . As in the zeroth approximation, we find  $x = \frac{1}{2}$  is always a solution, but that this gives a maximum value to  $\Delta_m g$  below some critical temperature, and a minimum value above this temperature. Below the critical temperature, there are two other solutions symmetrically disposed about  $x = \frac{1}{2}$ .

The differentiation of eqn. (24.12) involves some rather lengthy algebra because  $\beta$  is a function of  $x$  (eqn. (24.9)). Equating the first differential coefficient to zero, we obtain

$$\ln \frac{x}{1-x} + \frac{z}{2} \ln \frac{(\beta-1+2x)(1-x)}{(\beta+1-2x)x} = 0$$

or

$$\left\{ \frac{x}{1-x} \right\}^{(z-2)/z} = \frac{\beta-1+2x}{\beta+1-2x} = C_5. \quad (24.13)$$

This is the equivalent of eqn. (23.7) of the zeroth approximation, and the roots give the equilibrium compositions of the coexisting phases at any temperature. It is possible to eliminate  $\beta$  from (24.13) and thus obtain a more satisfactory expression for the solubility limits. By manipulation of (25.13)

$$\beta^2 - (1-2x)^2 = 4C_5(1-2x)^2/(1-C_5)^2$$

and eliminating  $\beta$  using (24.9)

$$\exp(2\mathcal{E}/kT) = C_5(1-2x)^2/x(1-x)(1-C_5)^2$$

or

$$\exp(\mathcal{E}/kT) = \frac{1-x/(1-x)}{[x/(1-x)]^{1/z} - [x/1-x]^{1-1/z}}. \quad (24.14)$$

The equilibrium conditions are thus given in terms of the ratio  $x/(1-x)$ . The solutions are symmetrical; if one root is  $x_1$ , the other is  $1-x_1$ . This is a necessary consequence of the nearest neighbour model.

At the critical temperature, the two roots coincide at  $x = \frac{1}{2}$ . Substituting  $x = \frac{1}{2}$  into eqn. (24.14) gives an indeterminate equation, so we write  $x/(1-x) = 1 + \delta$ , and by expanding in terms of  $\delta$  and letting  $\delta \rightarrow 0$ ,

$$\exp(\mathcal{E}/kT_c) = z/(z-2). \quad (24.15)$$

For a f.c.c. structure,  $z = 12$  and

$$T_c = \frac{\mathcal{E}}{k} \ln 1.2 = z\mathcal{E}/2.19k,$$

whilst for a b.c.c. structure

$$T_c = \frac{\mathcal{E}}{k} \ln 1.33 = z\mathcal{E}/2.3k.$$

These values may be compared with eqn. (23.10) for the zeroth approximation. If we let  $z \rightarrow \infty$  in eqn. (24.15), we re-obtain the value  $T_c = z\mathcal{E}/2k$ . Guggenheim has emphasized that this is a general result, and any formula of the quasi-chemical method in the first approximation can be converted into the corresponding zeroth order approximation by letting  $z \rightarrow \infty$ .

Equation (24.11) is a much more satisfactory expression for the free energy of a solid solution than is (23.5), especially as it includes a temperature-dependent energy term. In practical applications, however, eqns. (23.7) and (23.10) are usually used in preference to (24.14) and (24.15) because of their greater simplicity. In Fig. 6.9, the predicted solubility

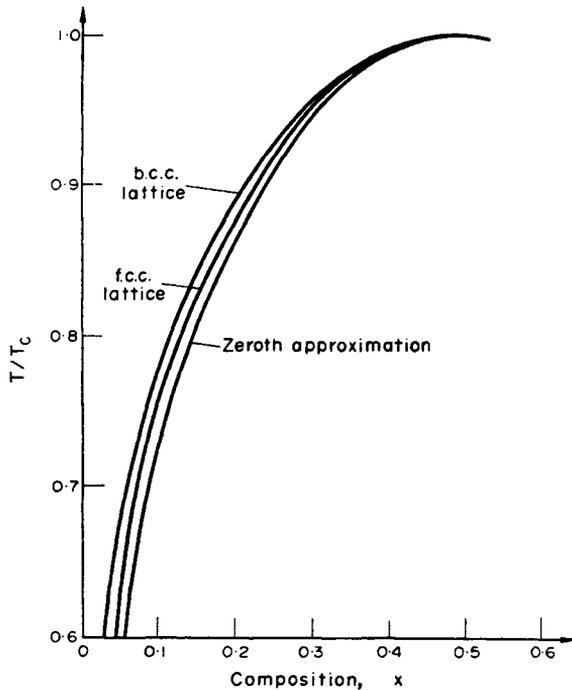


FIG. 6.9. Solubility limits according to the zeroth and first approximations of the quasi-chemical theory.

limits for the first and zeroth approximations are compared by plotting them as a function of  $T/T_c$ . The curve for the first approximation is drawn for both the f.c.c. and the b.c.c. structure. We have assumed identical values of  $T_c$  rather than of  $\mathcal{E}$ , since this most clearly reveals the shape of the curves. In practice,  $T_c$  is measured and  $\mathcal{E}$  is the disposable parameter.

The basic assumption of the method of calculation used in this section is the non-interference of atom pairs. This is a very artificial hypothesis since each atom belongs simultaneously to several pairs. Higher approximations may be obtained by considering the non-interference of interacting groups of larger numbers of sites. All such treatments may be classed as examples of the "cluster variation" method (Kurata, *et al.*, 1953), and the three-dimensional partition function of the Ising model is found approximately by a procedure similar to that outlined above for pairs of sites. A small group of lattice sites is chosen as the basic unit, the energy of this cluster is computed exactly for each way in which the sites may be occupied, and the number of microscopic configurations of the crystal is calculated approximately. The inconsistencies in the calculation of the number of configurations decrease as the size of the basic cluster increases, but the mathematics become correspondingly more complex. Most calculations have been made with triangular clusters of three sites, or with tetrahedral clusters of four sites; in some problems, involving superlattice formation, the choice of the basic cluster is governed by the crystallography in these higher approximations.

We shall not consider further the mathematical development of the Ising model; although the higher approximations lead to better solutions, they are mainly of value in order-disorder problems, where the use of pair interactions sometimes leads to difficulties (see p. 224). The results obtained do not generally differ greatly from those of the zeroth and first approximations; if systems of three sites are considered, for example, the critical temperature is given by

$$T_c = \Xi/k \ln\{(z+1)/(z-3)\}^{1/2},$$

and for a f.c.c. structure, this gives  $T_c = z\Xi/2.21k$ . Table V gives some comparisons of the predictions of the various cluster approximations; it is abstracted from more complete data given by Guggenheim (1952). The value of  $z$  in the cluster approximations is taken to be 12. The values of the excess entropy are negative and very small.

TABLE V. SUMMARY OF PREDICTIONS OF QUASI-CHEMICAL THEORY FOR THERMODYNAMIC FUNCTIONS AT  $T = T_c$

x	Function	Approximation			
		Zeroth	Pairs	Triplets	Quadruplets
0.1	$-\Delta_m g/kT$	0.15	0.13	0.13	0.13
0.3		0.19	0.17	0.17	0.16
0.5		0.19	0.17	0.17	0.17
0.1	$\Delta_m h/kT$	0.18	0.19	0.19	0.19
0.3		0.42	0.42	0.42	0.42
0.5		0.5000	0.50	0.50	0.50
0.1	$-\Delta_c s/k$	0	0.014	0.014	0.015
0.3		0	0.018	0.021	0.023
0.5		0	0.025	0.028	0.033
—	$z\Xi/kT_c$	2	2.19	2.21	2.23

We have already noted that exact solutions of the Ising model have been obtained in two dimensions. For a square lattice, the closed solution is

$$\sinh(z\mathcal{E}/4kT_c) = 1 \quad (24.16)$$

which gives

$$T_c = z\mathcal{E}/3.526k.$$

This formula has been used in Chapter 5, Section 19. Exact solutions in three dimensions have not been found in closed form, but quite early in the development of the theory Kirkwood pointed out that an exact solution for the partition function may be written as a power series in  $\mathcal{E}/kT$ . Unfortunately, this series converges very slowly, and since we are more interested in obtaining a reasonable physical model than in an expression for  $T_c$ , we shall not investigate this further.

In all the above derivations we have assumed that  $\mathcal{E}$  is a constant energy. The non-dependence of  $\mathcal{E}$  on the surrounding configuration of atoms is a fundamental postulate of the quasi-chemical theory, but Guggenheim (1948) has pointed out that there is no need to assume that  $\mathcal{E}$  is independent of temperature. An alternative approach is to regard  $\mathcal{E}$  as a quantity defined by eqn. (24.7). If we do this,  $2\mathcal{E}$  is the free energy increase when an  $A-A$  and a  $B-B$  pair are destroyed to produce two  $A-B$  pairs. This co-operative free energy may now include thermal entropy terms, insofar as these are affected by the interchange, and the restriction to solutions obeying the Kopp-Neumann rule has been removed. Since  $\mathcal{E}$  is now a function of temperature, we may write formally

$$\mathcal{E}' = \mathcal{E} - T(d\mathcal{E}/dT), \quad (24.17)$$

where  $\mathcal{E}'$  corresponds to our old definition of  $\mathcal{E}$  as an interchange energy. The equations for the thermodynamic quantities at any one temperature still contain only one parameter, which may be fixed empirically. On the other hand, the equations for the variation of solubility with temperature are no longer valid, and we cannot deduce  $\mathcal{E}$  or  $\mathcal{E}'$  from experimental data of this kind unless some assumption is made about  $d\mathcal{E}'/dT$ . Guggenheim suggested that the simplest assumption,  $\mathcal{E}'$  independent of temperature, may be adequate.

We now turn to a brief comparison of the predictions of the quasi-chemical theory with experiment. The comparison may be made in various ways; for example, values of  $\mathcal{E}$  can be determined from independently measured quantities and examined for consistency. If the theory is applicable,  $\mathcal{E}$  should also be independent of composition, and the composition dependence of experimentally measured activity coefficients and pair probability functions will show whether or not this is so. Detailed tests of this kind are possible in some assemblies, but are not needed in others, where the signs or the magnitudes of the excess thermodynamic quantities are sufficient to establish that the theory is unsatisfactory.

We note first of all that the formation of a metallic solid solution from its constituents produces fairly small changes in the thermodynamic functions, since the atomic interactions are not greatly changed. Thus, the experimentally measured excess functions are small; the heat of mixing per atom,  $\Delta_m g$ , usually lies within the limits  $\pm kT$ , and the excess entropy of mixing per atom,  $\Delta_m s$ , usually lies within the limits  $\pm k$ . Nevertheless, the magnitude of  $\Delta_m s$  provides one of the most convenient tests of the theory, and especially of the assumption

that the thermal entropy may be neglected. According to the quasi-chemical theory,  $\Delta_e s$  is always negative, and it is readily seen that  $|\Delta_e s/k|$  is unlikely to exceed 0.05 (see Table V). Data collected by Oriani (1959) for a large number of liquid and solid solutions (mainly at  $x = \frac{1}{2}$ ) show that this is rarely true. The experimental values range from  $\Delta_e s/k \approx -1$  for liquid  $Mg_{1/2} Bi_{1/2}$  to  $\Delta_e s/k \approx +0.7$  for solid  $Au_{1/2} Ni_{1/2}$ .

A related test of the assumption about thermal entropy is to consider the equilibrium solubility curves for dilute solid solutions, instead of measured thermodynamic functions. This has been done by Freedman and Nowick (1958) using eqn. (22.27). The partial excess entropies obtained in this way were all positive (with one doubtful exception) and of appreciable magnitude; this is true for assemblies giving  $\Delta_e s_B$  as well as for assemblies in which  $\Delta_e s_B$  (related to  $\Delta_e s_B$  by eqn. (22.26)) is obtained from the equilibrium diagram.

It follows from these results that the assumption of zero thermal entropy of mixing is a serious defect of the theory in most assemblies, and is more important than the error in the configurational term. This means that a theory with  $\bar{E}$  constant is not good enough, and the predictions of the model in respect of temperature variation are inadequate. However, if  $\bar{E}$  is allowed to vary with temperature, the quasi-chemical theory may still be a reasonable representation of the situation in some alloys at constant temperature, and it is, indeed, usually more successful in dealing with the heat of mixing than with the excess entropy.

We end this section by referring briefly to the attempts which have been made to develop better statistical models of solutions. The first of these is the theory of conformal solutions (Longuet-Higgins, 1951), which is a generalization of the pair interaction model, but is free from any structural assumptions. This makes it particularly suitable for dealing with liquid solutions.<sup>†</sup> A conformal solution is defined by certain assumptions about the form of the potential energy of the assembly, expressed as the sum of a number of bimolecular terms. These assumptions, which derive essentially from a law of corresponding states, are that the interaction energies between  $A-A$ ,  $B-B$ , and  $A-B$  pairs depend only on the nature and separation of the two atoms or molecules considered, and have the same functional dependence on the separation for all pairs. Thus the interaction energy of an  $A-A$  pair can be expressed in the form  $\bar{E}_{AA}(r) = f_{AA}\bar{E}_0(g_{AA}r)$ , with equivalent expressions for  $\bar{E}_{BB}(r)$  and  $\bar{E}_{AB}(r)$ . The factors  $f_{AA}$ ,  $f_{AB}$ ,  $g_{AA}$ , etc., are all constants, and  $-\bar{E}_0(r)$  is a universal function giving the potential energy of two atoms or molecules in a reference state as a function of their separation  $r$ . If the reference state is taken to be pure  $A$ , for example,  $f_{AA} = g_{AA} = 1$ .

In developing the theory, it is necessary to assume that  $1-f_{AA}$ ,  $1-g_{AA}$  are both small, so that all the forces have approximately the same magnitude, and it is also assumed that  $g_{AB} = \frac{1}{2}(g_{AA} + g_{BB})$ . The advantage of this approach is that the pair interaction energies vary with composition, since the mean separation is a function of composition, and the difficulties inherent in the usual quasi-chemical assumption of a temperature-independent  $\bar{E}$  are also avoided. The results of the theory are obtained in a series of successive approximations. The averaging procedure in the first approximation leads to a random distribution

<sup>†</sup> The quasi-chemical method is applicable in principle to liquids, and is often used for liquid solutions, but it does carry the implication of a "lattice model" of the liquid.

of the components, and is equivalent to the zeroth approximation of the theory of regular solutions; the higher approximations lead to statistical terms which are very difficult to evaluate (Brown and Longuet-Higgins, 1951).

Prigogine *et al.* (1957) have developed a theory of solutions which we may regard as based on the theory of conformal solutions and on the cell model. They use average interactions between a molecule or atom and its neighbours, and assume composition-dependent potentials of the Lennard-Jones (6-12) type. The interaction potentials are thus written  $-\Xi_{AA}(r) = -\Xi_{AA}^* \varphi(r/r_{AA}^*)$ , etc., where  $\Xi_{AA}^*$ , etc., are characteristic energies, and  $r_{AA}^*$ ,  $r_{AB}^*$ , etc., are characteristic lengths. The function  $\varphi$  is a universal function specifying the type of interaction assumed; for Lennard-Jones forces,  $\varphi(x) = x^{-12} - 2x^{-6}$ . The average potential model assumes random atomic distribution, and the expressions for the thermodynamic functions are obtained in terms of three parameters which are defined by the equations

$$\left. \begin{aligned} \varrho &= (r_{BB}^*/r_{AA}^*) - 1, \\ \delta &= (\Xi_{BB}^*/\Xi_{AA}^*) - 1, \\ \theta &= (\Xi_{AA}^* + \Xi_{BB}^* - 2\Xi_{AB}^*)/\Xi_{AA}^*. \end{aligned} \right\} \quad (24.18)$$

These parameters are essentially "reduced" quantities, and the theory gives expressions for  $\Delta_m h$  and  $\Delta_e s$  which can be simplified considerably when it is justifiable to neglect all products of  $\varrho$ ,  $\delta$ , and  $\theta$  and all powers of higher order than  $\delta$ ,  $\theta$ , and  $\varrho^2$  respectively. Under these conditions, which obviously correspond to those mentioned above in the theory of conformal solutions, we have

$$\left. \begin{aligned} \Delta_m h &\approx 1.435(\theta + 4.5\varrho^2)z\Xi_{AA}^*x(1-x), \\ \Delta_e s &\approx 78.2k\varrho^2x(1-x), \end{aligned} \right\} \quad (24.19)$$

where the factor 1.435 is a result of using Lennard-Jones forces for interactions between atoms which are not nearest neighbours; this factor is unity if only nearest neighbour interactions are considered. The results then reduce to the zeroth approximation of the quasi-chemical theory when  $\varrho = 0$ , and it will be seen that, according to this model, the positive excess entropy which is usually observed is attributable to the size disparity measured by  $\varrho$ .

Shimoji and Niwa (1957) have modified this theory by using generalized interatomic potentials of the Morse type in place of the 6-12 potential. They obtain an expression which is equivalent to Prigogine's equation for  $\Delta_m h$  if the Morse potential is chosen to correspond to Lennard-Jones forces, but they suggest that for liquid metal solutions, a more appropriate choice of potential gives

$$\Delta_m h \approx (\theta + 0.5\varrho^2)z\Xi_{AA}^*x(1-x). \quad (24.20)$$

The predictions of the average potential model have been compared with measured thermodynamic parameters by Oriani (1959). The procedure adopted was to assume  $r_{AB}^* = \frac{1}{2}(r_{AA}^* + r_{BB}^*)$  and to evaluate  $\theta$  from the measured  $\Delta_m g$ . The values of  $\Delta_m h$  and  $\Delta_e s$  were then calculated. The agreement found is rather poor, especially as the theory is only being used to decide how the excess free energy of mixing is distributed between the heat and excess entropy of mixing. The degree of agreement between the experimental and theoretical  $\Delta_m h$  is, moreover, little affected by whether or not the size disparity is taken into account.

The provisional conclusion is thus that the improved statistical theories are little better than the quasi-chemical theory in explaining the properties of solid metallic solutions. The main reason for this situation is that already emphasized on p. 111. Metals and metallic solutions do not obey a law of corresponding states, and the assumption of central force interactions of any type is a poor approximation.

## 25. MISFIT ENERGY IN SOLID SOLUTIONS: SUB-REGULAR SOLUTIONS

We have emphasized the restriction of the quasi-chemical theory to solutions in which the constituents have nearly equal atomic volumes. The failure of the theory in many cases in which large heats of mixing are associated with a considerable difference in the radii of the component atoms is frequently ascribed to "strain energy". This term is seldom precisely defined, but it is usually implied that the energy is a mechanical energy of long-range character (i.e. distributed throughout the crystal), and is somehow separable from any other (chemical) terms in the energy of mixing.

Theories of solid solution which utilize the notion of strain energy attempt to calculate the heat of mixing, or rather the partial excess free energy of mixing, from more fundamental parameters such as elastic constants. All treatments use the approximation of an elastic continuum to represent the solid, and then consider the mechanical strain energy when a solute atom is inserted into this continuum. The model used is one in which a sphere of specified radius is forced into a hole of smaller or larger radius, the surfaces welded together, and the body allowed to relax in its self-stressed state. Each solute atom is thus regarded as a centre of dilatation.

This model seems quite specific, but in fact the calculation has been applied in various ways. The most realistic physical assumption would seem to be to regard the atomic displacements as Hookean outside some limiting radius  $r_0$  from the centre of dilatation, and thus to calculate the strain energy in the region of crystal for which  $r > r_0$ . The total (free) energy of the solute atom, apart from the configurational term, would then be obtained by adding a "core energy" to the elastic energy; this is the approach which is normally used in dislocation theory (Chapter 7). A serious difficulty in the misfitting sphere model is that the elastic energy varies inversely as the cube of  $r_0$ , and so is very sensitive to the choice of this core radius. Different authors have put  $r_0$  equal to the radius of the solute atom itself, the radius of a cluster of twelve nearest neighbours, and the radius of the third nearest neighbour shell.

We should emphasize here that providing the processes envisaged above are isothermal, the calculation gives an estimate of the free energy<sup>†</sup> change associated with the introduction of the solute atom in a particular place, and not the internal energy or the enthalpy. Thus for an isolated solute atom, we are attempting to calculate  $\Delta_e g_B$ , and differentiation with

<sup>†</sup> Strictly, the elastic energy gives the change in the Helmholtz free energy, and the Gibbs free energy is obtained by adding the (negligible) work done against external forces in changing the volume of the assembly. The usual configurational term has, of course, to be added to the elastic energy to give the total free energy change associated with the solute atom.

respect to temperature will give  $\Delta_e s_B$ . This has been mentioned already in connection with Zener's conclusion about the sign of the excess (thermal) entropy of a point defect (p. 129). In some papers it is wrongly implied that the elastic energy is an internal energy.

The usual way of avoiding, or rather concealing, the difficulty of defining  $r_0$  is to apply elastic theory both to the surrounding matrix and to the solute atom and its immediate neighbours. Two energy terms are calculated, one being the work necessary to expand (contract) a hole in the matrix from an initial radius to a new final radius, and the other the work needed to compress (expand) a sphere from its initial radius to the final radius of the hole. The sum of these two contributions is then minimized to give the equilibrium final radius. The initial radii are commonly taken as those of a solvent and solute atom respectively, although a larger cluster may be used with equal justification.

Oriani (1956) has pointed out that this calculation is very unsatisfactory. The minimization procedure of the mechanical model has no real significance in the problem of determining the equilibrium "radius" of the solute atom. It is easy to see that the primary role in determining the distances of the first few coordination shells must be played by electronic interactions (including effects such as charge transfer and screening) with the immediate neighbours of the atom concerned. In a liquid solution, these interactions must entirely determine the "size" of the atom, since there are no long-range elastic stresses. For solid solutions, the same interactions must be present, and will still be the main factor determining the local configuration, but there will be a perturbation resulting from the imposition of lattice periodicity.

In a careful discussion of the significance of misfit energy, Oriani concludes that the term has no operational significance in liquid solutions, and any theoretical concept which might be invoked relates only to hypothetical situations which cannot be obtained physically. (For example, a theoretical definition of misfit energy might be the difference in energy of a real liquid solution and a hypothetical solution in which the components had the same ionic radii; this clearly has no operational significance and the quantity could never be measured.) The idea of size effects and misfit energy should thus be used very cautiously, if at all, for liquid solutions. For solid solutions it is possible to give an operational definition of misfit energy based on the changes produced by the imposition of long-range lattice periodicity. Only that part of the misfit energy outside some (unknown) radius  $r_0$  is susceptible to calculation by elastic theory, and might thus fairly be called strain energy.

These are severe criticisms of the elastic model, and we shall see later that they are to some extent supported by the experimental evidence. The model is capable of giving heats of mixing of the correct magnitude, and entropies of mixing of the correct sign and magnitude, but more detailed tests seem to show that  $\Delta_e g$  cannot really be ascribed to the elastic strain energy in the matrix. Nevertheless, we shall now give the elastic calculation in some detail. The reasons for this are twofold. First, similar calculations are still frequently used in many applications to both solid solutions and lattice defects, and as Eshelby (1956) has remarked, the limitations of this model are perhaps more immediately obvious than the equally serious defects of other approximations used in solid state theory. Second, and more important, we shall require the identical calculation in a later chapter when we consider the energy of a precipitate growing in a matrix. The minimization of the mechanical

strain energy then has much greater validity, and in fact the procedure was first used for this problem by Mott and Nabarro (1940).

In the present application, we consider a solid solution formed from two components of atomic volumes  $v_A^0$  and  $v_B^0$  in the respective pure states. Consider first the effect of introducing a single  $B$  atom into a lattice of pure  $A$ . According to the misfitting sphere model, we treat the  $B$  atom as an elastic sphere inserted into a hole of volume  $v_A^0$  in an isotropic elastic continuum.

Let the atomic radii of the atoms in the pure components be  $r_A^0$  and  $r_B^0 = (1 + \epsilon)r_A^0$ , where  $\epsilon$  may be positive or negative. We suppose first that the  $B$  atom is introduced into an infinite  $A$  crystal, where its volume becomes  $v_B$  and its effective radius is  $r_B = (1 + C_6\epsilon)r_A^0$ . Use an origin at the centre of the  $B$  atom. From the symmetry of the problem it follows that the displacements  $\mathbf{w}$  are all radial, and  $|\mathbf{w}|$  is a function only of the radius vector  $r$ . In polar coordinates, the strain components are

$$e_{rr} = \partial w / \partial r, \quad e_{\theta\theta} = e_{\phi\phi} = w/r,$$

and hence

$$\Delta = (\partial w / \partial r) + 2w/r, \quad \boldsymbol{\omega} = \text{curl } \mathbf{w} = 0. \quad (25.1)$$

The equations of equilibrium in the form (11.30) thus reduce to

$$(\lambda + 2\mu) \{ (\partial^2 w / \partial r^2) + (2/r) (\partial w / \partial r) - 2w/r^2 \} = 0,$$

and the general solution is

$$w = A_1 r + A_2 / r^2.$$

The displacements in any spherically symmetrical elastic problem must be of this type. In our infinite medium, we have to satisfy the boundary conditions

$$w(r_A^0) = C_6 \epsilon r_A^0, \quad w(\infty) = w(0) = 0, \quad (25.2)$$

if the initial state is taken to be the introduction of the compressed  $B$  atom into the hole of radius  $r_A^0$ . The displacements are thus

$$\left. \begin{aligned} w'_B &= C_6 \epsilon r & (r < r_A^0), \\ w_A &= C_6 \epsilon (r_A^0)^3 / r^2 & (r > r_A^0). \end{aligned} \right\} \quad (25.3)$$

The vector  $\mathbf{w}_A = C_6 \epsilon \{ (r_A^0)^3 / r^3 \} \mathbf{r}$  gives the displacements in the  $A$  crystal relative to the unstressed state. The chosen initial state for the  $B$  atom, however, is when it has been compressed or extended from radius  $(1 + \epsilon)r_A^0$  to  $r_A^0$ . The displacements  $w'_B$  refer to this state, and relative to the unstressed  $B$  sphere (before its introduction into the hole), we have the displacements

$$w_B = \epsilon(C_6 - 1)r,$$

so that within this sphere  $e_{rr} = e_{\theta\theta} = \Delta_B / 3 = (C_6 - 1)\epsilon$ . The state of stress is a uniform hydrostatic pressure (or tension) given by

$$p = 3K_B(C_6 - 1)\epsilon, \quad (25.4)$$

where  $K_B$  is the effective bulk modulus of the solute atom  $B$ . It is usual in this model to take  $K_B$  as the bulk modulus of the component  $B$ . This is admittedly a poor approximation because of the changes in the electronic state of a strongly deformed atom.

The introduction of the  $B$  atom leaves the whole assembly in a state of self-stress, and the stress components in the surrounding matrix are  $X_{rr}$  and  $X_{\theta\theta} = X_{\varphi\varphi}$ . From (25.1) and (25.3) we see that  $\Delta_A = 0$ , so that only shear strains are present, and the stresses may be written entirely in terms of the shear modulus  $\mu_A$  of  $A$ . Using (11.26) the stress components are

$$\left. \begin{aligned} X_{rr} &= 2\mu_A(\partial w/\partial r) = -4\mu_A C_6 \varepsilon(r_A^0)^3/r^3, \\ X_{\theta\theta} &= 2\mu_A w/r = 2\mu_A C_6 \varepsilon(r_A^0)^3/r^3. \end{aligned} \right\} \quad (25.5)$$

Since the matrix and inserted sphere are in equilibrium,  $X_{rr}$  must equal  $p$  (eqn. (25.4)) when  $r = r_B \approx r_A^0$ . This gives an expression for  $C_6$ :

$$C_6 = 3K_B/(3K_B + 4\mu_A). \quad (25.6)$$

If the  $B$  atom is incompressible ( $K_B = \infty$ ),  $C_6 = 1$  and  $v_B = v_B^0$ .

Let  $S$  be any closed surface in the assembly which totally encloses the  $B$  atom. An element  $dO$  of this surface, having unit vector normal  $\mathbf{n}$ , moves when the  $B$  atom is introduced so as to sweep out a volume  $\mathbf{w}_A \cdot \mathbf{n} dO$ . The volume within  $S$  thus increases by

$$\left. \begin{aligned} \Delta v^\infty &= C_6 \varepsilon(r_A^0)^3 \int_S (\mathbf{r} \cdot \mathbf{n}/r^3) dO \\ &= 4\pi C_6 \varepsilon(r_A^0)^3 \end{aligned} \right\} \quad (25.7)$$

since the integral is just the total solid angle subtended by  $S$  at the origin. We write the change of volume  $\Delta v^\infty$  to emphasize that this equation is only valid in an infinite crystal. Note that  $\Delta v^\infty$  is equal to the increase in volume of the hole, and is independent of  $S$ , as is necessary since there is no dilatation outside the  $B$  atom. If we write the difference in atomic volumes of the two pure components as  $\Delta v_{AB} = v_B^0 - v_A^0 = 4\pi \varepsilon(r_A^0)^3$ , we also have

$$\Delta v^\infty = C_6 \Delta v_{AB}. \quad (25.8)$$

Now consider what happens in a finite bounded crystal of  $A$ . The boundary conditions (25.2) are no longer applicable, and must be modified so that the external surface is free of traction. This may be accomplished by superimposing on the displacements (25.3) a second set of "image" displacements, caused by surface tractions which will just annul those given by (25.5). The work of Eshelby (1954, 1956) has shown that these image effects are often of surprising importance in the continuum theory of lattice defects.

Suppose the external surface is  $S$ . Then the image displacements are due to tractions  $-X_{ij}n_j$  distributed over  $S$ . If  $S$  is a sphere of radius  $R$ , the surface traction to be applied is a uniform hydrostatic tension (or pressure) of  $4\mu_A C_6 \varepsilon(r_A^0)^3/R^3$ . The corresponding image displacements represent a uniform dilatation of the large sphere provided we neglect the small perturbation caused by the different elastic constants of the enclosed  $B$  atom. This dilatation is

$$\Delta_A = 4\mu_A C_6 \varepsilon(r_A^0)^3/(K_A R^3), \quad (25.9)$$

and the increase in volume of the whole assembly is

$$\left. \begin{aligned} \Delta v^i &= 16\pi\mu_A C_6 \varepsilon (r_A^0)^3 / 3K_A \\ &= (4\mu_A / 3K_A) \Delta v^\infty. \end{aligned} \right\} \quad (25.10)$$

Since the dilatation varies as  $1/R^3$ ,  $\Delta v^i$  is independent of  $R$ . Thus for any sphere, the total change in volume is

$$\left. \begin{aligned} \Delta v &= \Delta v^\infty + \Delta v^i = 4\pi C_6 \varepsilon (r_A^0)^3 / C_6' \\ &= (C_6 / C_6') \Delta v_{AB}, \end{aligned} \right\} \quad (25.11)$$

where  $C_6' = 3K_A / (3K_A + 4\mu_A)$ .

If the bulk moduli of the matrix and the  $B$  atom are equal,  $C_6 = C_6'$  and  $\Delta v = \Delta v_{AB}$ . This illustrates the importance of including the image effects, since this result cannot be otherwise obtained.

A solution for the image displacements is possible only when the surface  $S$  has certain simple forms. However, it is readily proved (Eshelby, 1956) that the result for the change in volume is valid for any external surface. In a sphere with a uniform distribution of solute atoms,  $c_B$  per unit volume, it follows from symmetry that the shape of the assembly is unaltered, apart from small ripples in the surface, of magnitude determined by the mean separation of  $B$  atoms. The fractional change in the volume of the assembly will be

$$\Delta V / V = c_B (C_6 / C_6') \Delta v_{AB}. \quad (25.12)$$

Eshelby has shown that this result remains true for a body of arbitrary shape which contains a uniform distribution of defects.

Returning to the single  $B$  atom in otherwise pure  $A$ , we now calculate the strain energy. Within the compressed or extended  $B$  sphere, the strain energy density,  $\frac{1}{2}p\Delta$ , is constant, and the total strain energy is thus

$$W_B = 9v_B K_B (C_6 - 1)^2 \varepsilon^2 / 2. \quad (25.13)$$

If the atom is in an infinite matrix, the strain energy density at any point is

$$\frac{1}{2} X_{rr} e_{rr} + X_{\theta\theta} e_{\theta\theta} = 6\mu_A C_6^2 \varepsilon^2 (r_A^0)^6 / r^6,$$

and the total strain energy by integration is

$$W_A = 6\mu_A C_6^2 \varepsilon^2 v_B \quad (25.14)$$

to the first order in which all these equations are valid. The same result for  $W_A$  is obtained by considering the strain energy density in a finite medium and integrating from  $r_B$  to  $R$ .

The strain energy of the whole assembly is thus

$$W_s = W_A + W_B = 6\mu_A C_6 \varepsilon^2 v_B \quad (25.15)$$

using the result of (25.6). Since  $\varepsilon = \Delta v_{AB} / 3v_B$ , we may also write the strain energy in the form

$$W_s = 2\mu_A C_6 (\Delta v_{AB})^2 / 3v_B. \quad (25.16)$$

Note that the volume  $v_B$  in eqns. (25.13), (25.14), and (25.16) might equally well be written  $v_A^0$  or  $v_B^0$  to the order of approximation of linear elasticity theory, and it would perhaps be better to write these equations with some mean volume  $v$ .

We can now consider the properties of a solid solution containing  $N_A$   $A$  atoms and  $N_B$   $B$  atoms. If  $N_B$  is small, the volume of the assembly will be  $V + N_B \Delta v$ , where  $\Delta v$  is given by eqn. (25.11). The mean volume per atom is thus

$$\overline{v_{AB}} = v_A^0 + x \Delta v,$$

where  $x$  is the atomic concentration of  $B$  atoms. Substituting  $\Delta v = (C_6/C_6')(v_B^0 - v_A^0)$ , we find

$$\overline{v_{AB}} = (1-x)v_A^0 + xv_B^0 + x(C_6/C_6' - 1)(v_B^0 - v_A^0). \quad (25.17)$$

When  $K_A = K_B$ , this leads to an additive law for atomic volumes, or on the linear approximation to an additive law for atomic radii

$$\overline{r_{AB}} = (1-x)r_A^0 + xr_B^0 = (1 + \varepsilon x)r_A^0, \quad (25.18)$$

so that  $\varepsilon$  is the fractional rate of change of lattice constant  $\partial r/\partial x$ . Equation (25.18) is usually known as Vegard's law, and according to the elastic model the radii are larger or smaller than their Vegard's law values according to whether the last term in (25.17) is positive or negative. The sign of the deviation is thus dependent on the sign of

$$(K_B - K_A)(r_B^0 - r_A^0).$$

Friedel (1955) has shown that this rule accounts for observed qualitative deviations from the law (using X-ray measurements), although quantitative agreement, as expected, is not very good.

In deriving eqn. (25.17) we have assumed that  $\Delta v$  is a constant independent of  $x$ , and we have also neglected the variation of elastic constants with  $x$ . It might strictly be more logical to assume that the  $B$  atom is forced into a hole of size  $v_A^0(1 + \Delta_A)$ , where  $\Delta_A$  is the dilatation produced in the matrix by all the preceding  $B$  atoms. However, the misfitting sphere model is probably too uncertain in its details for this kind of modification. Eshelby has pointed out that there is no very convincing reason to take the volumes of the hole and misfitting sphere as  $v_A^0$  and  $v_B^0$  respectively; they might, for example, have been chosen so that their radii were equal to nearest neighbour distances in the pure components. Experimental results which show that Vegard's law is approximately valid in many alloys are the best justification for the assumption we have made.

We are now able to derive the strain energy of the solid solution. When the composition is  $x$ , the addition of a further  $B$  atom changes the energy first by the self-energy  $W_s$  of this atom, and, secondly, by the interaction energy of the new solute atom with all the preceding  $B$  atoms. This interaction energy is due to the presence of a hydrostatic tension or compression in the matrix (the image stress), and as this helps the required expansion or contraction of the hole, the total extra energy is less than  $W_s$ . Although the interaction energy of any two  $B$  atoms is negligible in comparison with  $W_s$ , the fact that each  $B$  atom interacts with

all the other  $B$  atoms means that the total interaction energy is comparable with the total self-energy.

When the concentration is  $c_B$ , the dilatation is  $c_B \Delta v^i = (x/v_A^0) \Delta v^i$ . The interaction energy when another  $B$  atom is added is thus  $K_A(x/v_A^0) \Delta v^i \Delta v = 2W_s(C_6/C_6')x \simeq 2W_s(K_B/K_A)x$ . The net change in energy on adding the  $B$  atom is thus

$$\delta W = W_s \{1 - 2(K_B/K_A)x\},$$

and this corresponds to a change of composition  $\delta x = 1/N$ . The total strain energy is thus

$$W = NW_s x \{1 - (K_B/K_A)x\} \quad (x \ll 1). \quad (25.19)$$

If chemical interaction terms are small, this strain energy is the atomic part of the free energy of solution, as pointed out on p. 198, and the total free energy of solution is obtained by adding the configurational term for random mixing to give

$$\Delta_m g = W_s x \{1 - (K_B/K_A)x\} + kT \{x \ln x + (1-x) \ln(1-x)\} \quad (x \ll 1). \quad (25.20)$$

Except for the factor  $(K_B/K_A)$ , this equation is identical in form with eqn. (23.5) for a regular solution. However, we have already emphasized that  $W$  is a free energy, so that the temperature variation of the elastic constants gives rise to an appreciable temperature variation of  $W$ , and hence to an entropy

$$\Delta_e s = -(\partial W / \partial T). \quad (25.21)$$

From eqn. (25.16) we see that the temperature variation of  $W$  is governed by the term  $\partial \mu / \partial T$ , and since this is always negative, the elastic model always gives a positive value to  $\Delta_e s$  (Zener, 1951). Friedel (1955) has compared the values of  $\Delta_e s$  and  $\Delta_m h$  obtained from eqns. (25.21) and (25.19) with experimental data on gold-nickel alloys, and obtained good agreement. In a survey of limiting solubility data for a number of binary alloys, using the method of analysis described on p. 200, Freedman and Nowick (1958) found that  $\Delta_e s$  is always positive, in accordance with the predictions of the model. The model actually permits calculation of  $\Delta_e s$ , but, as pointed out earlier, it is too sensitive to a choice of integration limits for this calculation to have much significance. The values obtained are certainly of the correct magnitude, but it is also reasonable to expect some proportionality between the observed  $\Delta_e s$  and that calculated from the size factor and temperature variation of the shear modulus. Freedman and Nowick showed that such a proportionality does not exist, and concluded that a quantitative interpretation of  $\Delta_e s$  cannot be given by the elastic model. Such an interpretation almost certainly requires detailed consideration of nearest neighbour interactions, as in Huntington's calculations for point defects (see p. 130). Oriani (1959) reaches the same conclusion, which he bases partly on the observed experimental correlation between heats of solution in the liquid and solid phases of alloys with a size disparity. The elastic model is not applicable to a liquid solution, so there should be no correlation if it is valid.

Within the limitations of the model, the equations derived above are strictly valid only for very dilute solutions. Some attempts have been made to treat more concentrated solutions by supposing that each atomic site is occupied by a positive or negative strain centre

acting in a mean lattice. We shall not describe these theories because of their complexity; some attention has been paid to the X-ray diffraction effects expected from a model of this type. The simplest treatment of more concentrated solutions comes from noting that eqn. (25.20) will be nearly identical for both *A*-rich and *B*-rich solutions if the components have nearly identical elastic properties, and as a first approximation may be applied to intermediate compositions, using mean values for the parameters. Comparison with the previous quasi-chemical treatment then shows that the critical temperature below which phase segregation occurs is given by

$$T_c = 3\bar{C}_8 \varepsilon^2 \bar{v} \bar{\mu} / k,$$

and if this is to be below the melting point of the alloy,

$$\varepsilon^2 < (kT^s / 3\bar{C}_8 \bar{v} \bar{\mu}).$$

Reasonable values of the constants give  $\varepsilon \sim 0.15$ , in agreement with the empirical Hume-Rothery rule that wide solid solution is possible only when the size factor  $\varepsilon$  is less than about 15%. In view of the reservations above, it is doubtful if this is anything other than a coincidence.

An early attempt at a semi-empirical equation for the strain energy was made by Lawson (1947). He used the expression for the self-energy of a *B* atom in an *A* lattice, but replaced  $\Delta v_{AB}$  by  $v_B^0 - v_{AB}$ . This means that the hole into which the *B* atom is forced is equal in volume to the mean volume per atom in the assembly. Lawson neglected interaction energies, and his equation for the free energy may be expressed in the form

$$\Delta_m g = A_1 x(1-x)^2 + A_2 x^2(1-x) + kT \{x \ln x + (1-x) \ln(1-x)\}. \quad (25.22)$$

The assumptions made in deriving the equation are too sweeping for the expressions for  $A_1$  and  $A_2$  to have any real value, but the equation itself may be useful as a representation of the properties of some real solutions. Its real advantage as an empirical equation is that it effectively replaces the constant  $\bar{E}$  of the quasi-chemical theory by a composition dependent exchange energy, and the assumed independence of the quantities  $\bar{E}_{AA}$ ,  $\bar{E}_{AB}$ , etc. of composition is a major weakness of the quasi-chemical theory.

The formal properties of eqn. (25.22) have been investigated by Hardy (1953), who used it to define a "sub-regular solution". For the present, we assume that  $A_1$  and  $A_2$  are independent of temperature and composition, and the chemical potentials are then

$$g_A - g_A^0 = x^2(2A_1 - A_2) + 2x^3(A_2 - A_1) + kT \ln(1-x). \quad (25.23)$$

The compositions of the solubility curve at any temperature are found by differentiating (25.22) and applying the condition of eqn. (22.5). This gives

$$kT \ln \left\{ \frac{x_2(1-x_1)}{x_1(1-x_2)} \right\} + (x_2 - x_1) \{2(A_2 - 2A_1) + 3(A_1 - A_2)(x_1 + x_2)\} = 0, \quad (25.24)$$

where, as before,  $x_1$ ,  $x_2$  are the limits of solubility of *B* in *A* and *A* in *B* respectively. The equation of the spinodal is similarly found to be

$$2(A_2 - 2A_1) + 6x(A_1 - A_2) + kT/x(1-x) = 0, \quad (25.25)$$

and at the critical point

$$\partial^3 \Delta_m g / \partial x^3 = 6(A_1 - A_2) + (2x - 1) kT / x^2(1 - x)^2 = 0. \quad (25.26)$$

From eqns. (25.25) and (25.26) the critical temperature  $T_c$  is related to the critical composition at which the solubility gap closes,  $x_c$ , by the equations

$$\left. \begin{aligned} A_1 &= kT_c(-9x_c^2 + 8x_c - 1)/6x_c^2(1 - x_c)^2. \\ A_2 &= kT_c(-9x_c^2 + 10x_c - 2)/6x_c^2(1 - x_c)^2. \end{aligned} \right\} \quad (25.27)$$

From eqns. (25.27) we see that if  $A_1 = A_2$ ,  $x_c = \frac{1}{2}$  and  $T_c = A_1/2k$ . The model is then formally equivalent to the quasi-chemical model. As  $A_1 - A_2$  increases, the composition of the maximum in the solubility gap moves further away from the equi-atomic composition, and the asymmetry of the solubility gap increases.

Instead of using eqn. (22.5), we could have obtained (25.24) from the equivalent thermodynamic conditions

$$g_A(x_1) = g_A(x_2), \quad g_B(x_1) = g_B(x_2).$$

If we multiply the first of these equations by  $2 - (x_1 + x_2)$  and the second by  $(x_1 + x_2)$  and then add, we obtain

$$(2 - x_1 - x_2) kT \ln\{(1 - x_1)/(1 - x_2)\} + (x_1 + x_2) kT \ln(x_1/x_2) = -(A_1 - A_2)(x_2 - x_1)^3. \quad (25.28)$$

In a regular solution, both sides of this equation are zero. Hardy used (25.28) to test the sub-regular solution model for a number of alloy assemblies. Provided  $(A_1 - A_2)$  is independent of temperature, a straight line may be obtained by plotting the left-hand side of the equation (obtained from the experimental equilibrium diagram) against  $(x_2 - x_1)^3$ . In this way, the alloys Ag-Cu, Ag-Pt, Al-Zn, and Au-Pt were shown to behave approximately as sub-regular solutions. The simple model could not be fitted to other alloys, e.g. Au-Fe, Au-Co, and Au-Ni, where  $A_1 - A_2$  is not constant and  $A_1$  and  $A_2$  are apparently functions of composition. In the alloys which may be represented as sub-regular solutions,  $A_1$  and  $A_2$  may increase with temperature, although the difference  $A_1 - A_2$  is nearly independent of temperature.

The theory of sub-regular solutions does not rest on any firm basis, and has very limited utility. It supplies an empirical equation for the solubility curve in terms of a single parameter  $(A_1 - A_2)$  (eqn. (25.28)), and is thus useful in representing the properties of alloys with equilibrium diagrams which are not symmetrical about the equi-atomic composition.

## 26. ORDERED STRUCTURES IN ALLOYS

In the previous sections we have concentrated mainly on deriving the equilibrium conditions for an assembly with a positive heat of mixing, i.e. an assembly in which  $\bar{E}$  is positive according to the nearest neighbour model. As already noted, a negative  $\Delta_m h$  (negative  $\bar{E}$ ) implies that atoms of opposite kinds attract each other. At low temperatures, each  $A$  atom will thus surround itself with as many  $B$  atoms as possible; such a structure is called a super-lattice. A completely ordered structure of this kind resembles a chemical compound, and

is possible only when the atomic fractions of the different components are small integral numbers.

The term superlattice occasionally causes some confusion. As described on p. 171, we do not need to distinguish between the different kinds of atom in a substantially disordered solid solution, since they occupy the available sites approximately at random. When the structure is ordered, the atoms segregate in such a way that the atoms of one kind occupy one or more sets of sites, and atoms of another kind occupy different sets of sites. A distinction must then be made between vectors specifying the positions of *A* atoms and those specifying the positions of *B* atoms, and the size of the primitive unit cell of the Bravais lattice has to be increased. This is the origin of the term superlattice. The conventional unit cell may remain the same size, as happens, for example, in the simplest type of superlattice, which is formed from a disordered b.c.c. structure. If there are equal numbers of *A* and *B* atoms, they may arrange themselves so that all the *A* atoms are at cube corners and all the *B* atoms at cube centres. The superlattice unit cell is the same as the cubic unit cell of the b.c.c. structure, but the structure is now simple cubic because of the non-equivalence of the corner and centre atoms. The cubic unit cell is the smallest possible for the superlattice structure, but the disordered structure has a smaller rhombohedral unit cell. In other superlattices, more complex changes take place, and the formation of the superlattice may be accompanied by a lowering of lattice symmetry.

It would be inappropriate in this book to discuss the structural features of all the known superlattices, but it is convenient to summarize the main types. In binary alloys, most superlattices have one of five structures, two being derived from each of the common cubic structures and one from the h.c.p. structure. In addition, there are a number of superlattice structures of which only one example is known, and some structures which have large unit cells, corresponding to a modification of one of the basic types.

The superlattices derived from the f.c.c. structure (*A1*) are known as  $L1_0$  and  $L1_2$  in the Strukturbericht notation. The unit cell of each contains the four atoms found in the cubic unit cell of the disordered *A1* structure, but  $L1_0$  has tetragonal symmetry. In this structure, which occurs at equi-atomic positions, the *A* atoms are at points [000] and  $[\frac{1}{2} \frac{1}{2} 0]$  of the unit cell, and the *B* atoms at points  $[\frac{1}{2} 0 \frac{1}{2}]$  and  $[0 \frac{1}{2} \frac{1}{2}]$ . The structure consists of alternate layers of *A* and *B* atoms parallel to the (001) planes. The attraction between *A* and *B* atoms results in slightly smaller interatomic distances between nearest neighbours in adjacent layers, so the structure is tetragonal with *c/a* slightly smaller than unity. Each atom has four nearest neighbours of its own type in the same (001) layer, and eight nearest neighbours of opposite type in the two adjacent (001) layers. This contrasts with the completely disordered structure, where each atom on the average has six like and six unlike nearest neighbours.

The  $L1_2$  structure corresponds to the ideal composition  $A_3B$ . The *B* atoms are in the [000] positions, and the *A* atoms in the remaining positions of the conventional unit cell of the f.c.c. structure. In the superlattice, the *B* atoms each have twelve unlike nearest neighbours, compared with an average of three like and nine unlike nearest neighbours in a random f.c.c. solid solution. There is also a single known example of the  $L1_1$  structure, in which alternate (111) planes are composed entirely of *A* and *B* atoms respectively.

The simplest superlattice ( $B2$  or  $L2_0$ ) derived from the b.c.c. ( $A2$ ) structure is found in equi-atomic alloys, and was described above. The symmetry is simple cubic, and the unit cell contains the two atoms of the conventional b.c.c. unit cell. The  $A$  atoms are at the corners  $[000]$  and the  $B$  atoms at the body centres  $[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$ , or vice versa; this is sometimes called the caesium chloride structure. Each atom has eight unlike nearest neighbours in the disordered state compared with an average of four like and four unlike neighbours in the completely disordered state. Another superlattice derived from the b.c.c. structure is more complex; it is known as the  $D0_3$  type. This structure has a cubic unit cell formed from eight conventional b.c.c. unit cells, and thus containing sixteen atoms; the ideal composition is  $A_3B$ . The corner positions of the small cubic cells are occupied by equal numbers of the atoms of each kind, each set being arranged on tetrahedral groups of sites. The body centred positions of the small unit cells are occupied entirely by  $A$  atoms. Each  $B$  atom has eight unlike nearest neighbours in the superlattice, compared with an average of two like and six unlike nearest neighbours in the substantially disordered solid solution.

From the h.c.p. structure ( $A3$ ), a superlattice is made by stacking together four unit cells of two atoms to give a larger unit cell of sides  $2a, 2a, c$  containing eight atoms. The composition is again  $A_3B$ , and each close packed layer contains three times as many  $A$  atoms as  $B$  atoms. The  $B$  atoms form a hexagonal network of side  $2a$ , and the  $A$  atoms occupy the remaining sites of the hexagonal network of side  $a$ . Each  $B$  atom is surrounded by twelve  $A$  atoms in the superlattice compared with an average of three like and nine unlike atoms in the random solution.

The relations we have just described are summarized in Table VI, which also lists some of the known binary assemblies having superlattices of these kinds. In some cases, the

TABLE VI. COMMON SUPERLATTICE STRUCTURES

Disordered structure	Superlattice type	Composition	Atom positions	Examples
f.c.c.	$L1_0$ (tetragonal)	$AB$	$2A$ in $(000; \frac{1}{2}, \frac{1}{2}, 0)$ $2B$ in $(\frac{1}{2}, 0, \frac{1}{2}; 0, \frac{1}{2}, \frac{1}{2})$	AuCu, CoPt, MgIn MnNi, NiPt, FePd, FePt
f.c.c.	$L1_2$ (cubic)	$A_3B$	$3A$ in $(0\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, 0, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 0)$ $1B$ in $(000)$	$Cu_3Au, Au_3Cu,$ $Pt_3Co, Fe_3Pt, Pt_3Fe,$ $Cu_3Pt, Ni_3Mn, etc.$
b.c.c.	$B2 (L2_0)$ (cubic)	$AB$	$1A$ in $(000)$ $1B$ in $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$CuZn, CuPd, AgCd,$ $AgZn, CoFe$
b.c.c.	$D0_3$ (cubic, face-centred)	$A_3B$	$(000; 0\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \frac{1}{2}, 0, \frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 0)$ $+ 4B$ in $(000)$ $4A$ in $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ $8A$ in $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}; \frac{3}{4}, \frac{3}{4}, \frac{3}{4})$	$Fe_3Al, Fe_3Si,$ $Mg_3Li, Cu_3Al$
h.c.p.	$D0_{19}$ (hexagonal)	$A_3B$	$6A$ in $(\frac{1}{2}, 0, 0; 0, \frac{1}{2}, 0; \frac{1}{2}, \frac{1}{2}, 0);$ $\frac{1}{6}, \frac{1}{3}, \frac{1}{2}; \frac{1}{6}, \frac{5}{6}, \frac{1}{2}; \frac{2}{3}, \frac{5}{6}, \frac{1}{2})$ $2B$ in $(000, \frac{2}{3}, \frac{1}{3}, \frac{1}{2})$	$Ag_3In, Mn_3Ge,$ $Mg_3Cd, Cd_3Mg,$ $Ni_3Sn$

nearest neighbour model with negative  $\bar{\mathcal{E}}$  predicts that for a given crystal structure and composition the lowest energy state will correspond to an observed superlattice type. However, this is not always valid, and the model may be ambiguous; for example, the energy of a b.c.c. solution is reduced by the formation of a phase mixture of the  $L2_0$  superlattice and pure  $A$ , so that all  $B$  atoms have only  $A$  atoms as nearest neighbours. For an  $AB_3$  alloy it follows that the energy of this phase mixture is identical with that of a  $DO_3$  superlattice, so that the nearest neighbour model is not able to predict which is the equilibrium state. It then becomes necessary to consider higher neighbour interactions; this procedure is not so artificial as first appears since, as previously noted, the interchange energies can be given operational definitions (Clapp and Moss, 1968).

Several early calculations of the lowest energy states with second and sometimes third nearest neighbour interactions taken into consideration were restricted either to stoichiometric compositions or to one-phase states. A systematic investigation for cubic alloys with first and second neighbour interactions was first made by Richards and Cahn (1971) and was supplemented by Allen and Cahn (1972). The results are expressed in terms of the ratio  $\xi = \bar{\mathcal{E}}_2/\bar{\mathcal{E}}_1$  (see eqn. (23.1a)) which is unrestricted in sign and magnitude; Richards and Cahn considered only  $\bar{\mathcal{E}}_1 < 0$ , but Allen and Cahn also treated  $\bar{\mathcal{E}}_1 > 0$ . Because of symmetry, it is only necessary to consider explicitly the composition range  $0 \leq x \leq \frac{1}{2}$ .

The internal energy of any configuration may now be expressed in an obvious extension of eqn. (24.1) as

$$\begin{aligned} \Delta_m U &= z_1 N_{AB,1} \bar{\mathcal{E}}_1 + z_2 N_{AB,2} \bar{\mathcal{E}}_2 \} \\ &= \frac{1}{2} N \bar{\mathcal{E}}_1 [z_{AB,1} + z_{AB,2} \xi], \} \end{aligned} \quad (26.1)$$

where  $z_{AB,i}$  now denotes the average number per atom of  $i$ th neighbours of opposite type. Thus for  $\bar{\mathcal{E}}_1$  negative the equilibrium state at 0 K (ground state) is found by maximizing the quantity in square brackets, whilst for  $\bar{\mathcal{E}}_1$  positive the energy will be minimized by minimizing this quantity. There are limits on  $z_{AB,1}$  and  $z_{AB,2}$ , given for  $0 \leq x \leq \frac{1}{2}$  by

$$0 \leq z_{AB,i} \leq 2z_i x. \quad (26.2)$$

The lower limit corresponds to all  $B$  atoms having only  $B$  atoms for  $i$ th neighbours; one way of attaining this is to postulate a two-phase mixture of the two components in which  $z_{AB,i} = 0$  for all  $i$ . At the upper limit, each  $B$  atom has only  $A$  atoms for  $i$ th neighbours, assuming  $B$  to be the minority component.

Clearly, if  $z_{AB,1}$  and  $z_{AB,2}$  can have the upper or lower limiting values (depending on the signs of  $\bar{\mathcal{E}}_1$  and  $\bar{\mathcal{E}}_2$ ) which will minimize  $\Delta_m U$ , a ground state has been found. Thus if both  $\bar{\mathcal{E}}_1$  and  $\bar{\mathcal{E}}_2$  are positive, minimum energy is given by the phase mixture ( $\Delta_m U = 0$ ) as already concluded. However, the crystal structure does not always allow the upper limits of (26.2) to be attained; for example  $z_1 = 12$  for a f.c.c. structure, but the maximum value of  $z_{AB,1}$  for an equi-atomic superlattice is 8. This arises because in this structure two nearest neighbours of a given atom may also be nearest neighbours of each other, and the upper limit of (26.2) is in fact only possible for  $0 \leq x \leq 0.25$ . In addition to this difficulty,  $z_{AB,1}$  and  $z_{AB,2}$  may not be independent of each other, so that it may be impossible to

find a configuration which produces the appropriate limits simultaneously, even though both may be individually obtained.

In cases where  $B-B$  bonds cannot be avoided for negative  $\bar{\mathcal{E}}_i$ , or  $A-B$  bonds cannot be avoided for positive  $\bar{\mathcal{E}}_i$ , the equilibrium structure(s) can only be found by calculating the energies of various possible configurations. Richards and Cahn used an empirical procedure of examining superlattices with translational vectors that coincide as much as possible with energetically favourable distances for like atoms. The results of this procedure seem intuitively to be correct, but no proof could be given that all possible low energy configurations had been considered. Allen and Cahn used a more rigorous technique, the cluster method, that includes consideration of all possible structures but also sometimes includes cluster combinations which represent impossible "structures".

In the cluster method, a motif of  $M$  adjacent lattice points (atoms) is defined by a specific three-dimensional circuit and the  $2^M$  clusters which correspond to the various occupancies of these sites by  $A$  and  $B$  atoms are enumerated. An energy can now be assigned to each cluster and the total energy of any arrangement of  $A$  and  $B$  atoms on a lattice can then be expressed as a linear sum of each such energy multiplied by the number of clusters of that type. The problem of minimizing this energy function subject to the constraints imposed by the overall composition can be solved by a mathematical technique known as linear programming and yields as a result the fractional numbers of clusters of each type in the whole arrangement. This gives an absolute minimum to the energy since all possible structures (single and multiphase) are represented. However, whilst the fractions of each cluster in any arrangement can always be specified, it is not always possible to construct an arrangement corresponding to specified fractions. Thus the numbers given by the technique may represent an "imaginary" structure, and the method then gives only a lower limit to the energy unless a true minimum energy structure can be found by selection of a different motif.

In the case of the b.c.c. structure, the results obtained by the cluster method for negative  $\bar{\mathcal{E}}_1$  coincide with those found by Richards and Cahn, and thus constitute a proof that their ground state diagram is correct. The equilibrium phases at 0 K are shown schematically in Fig. 6. 10 in which the phase fields are plotted as functions of  $\xi = \bar{\mathcal{E}}_2/\bar{\mathcal{E}}_1$  and of atomic per cent of  $B$ . The only structure not listed in Table VI is the  $B32$  (NaTi) type. For  $\xi$  negative a two-phase state is stable over the whole composition range, but for  $\xi$  positive the situation is more complex. For  $0 \leq \xi \leq \frac{2}{3}$  and 25–50 atomic %  $B$ , for example the same energy is obtained either by a phase mixture of the two stoichiometric phases,  $DO_3$  and  $B2$ , or by a gradual change in the occupancy of some of the sites. In the latter case, there is a continuous single phase region which includes the stoichiometric compositions. This degeneracy in the ground state is presumably a property of the particular model, and would disappear in a better approximation. Richards and Cahn point out that the configurational entropies of the degenerate states are not the same, and this will then determine the equilibrium state at finite low temperatures. On this basis, the fields of type  $DO_3 \rightarrow B2$  are perhaps better regarded as single phase.

It is particularly noteworthy that the ground state is multiply degenerate in the nearest neighbour model ( $\xi = 0$ ), except for  $x = \frac{1}{2}$ . At compositions near  $x = \frac{1}{4}$ , for example, it is possible for  $z_{AB,1}$  to have its maximum value of 8 (i.e. for there to be no  $B-B$  nearest

neighbour bonds) in a configuration consisting of (i)  $A+B_2$ , (ii)  $DO_3$ , (iii) single phase non-stoichiometric  $B_2$ , or (iv) an almost random structure. When second neighbour interactions are taken into account, this degeneracy disappears and either (i) or (ii) becomes stable even when  $|\xi|$  is very small. This is thus a strong argument for the inclusion of second neighbour interactions in some special situations. There are also degeneracies for  $\xi = \frac{2}{3}$  and  $0.25 \leq x \leq 0.5$ .

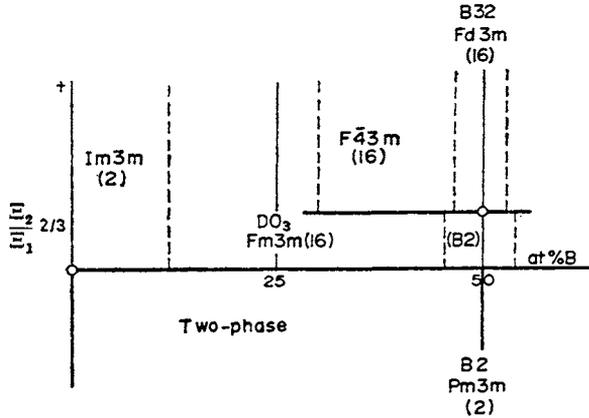


FIG. 6.10. Body-centred cubic ground state diagram showing the state of lowest energy as a function of composition and of the ratio of second to first neighbour interaction energies. Superlattices are identified by Strukturbericht and space-group symbols, with the number of atoms in the unit cell in brackets. Heavy lines show the limits of two-phase fields, and thin lines mark stoichiometric compositions which may be extended into single-phase fields as indicated by the broken lines (after Richards and Cahn, 1971).

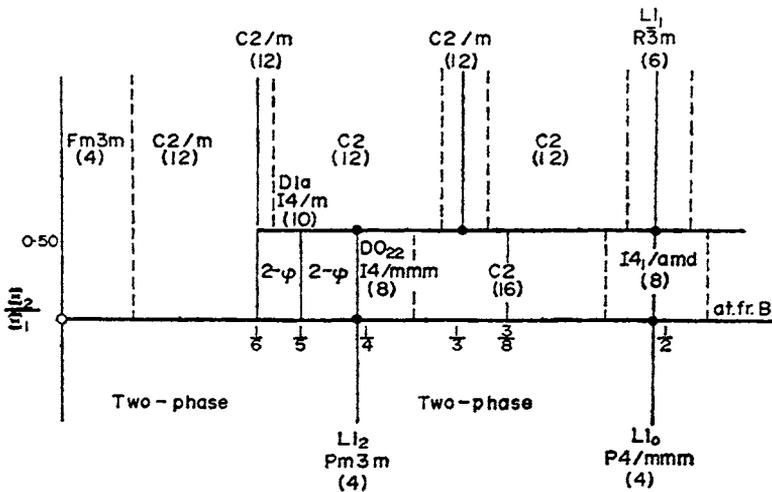


FIG. 6.11. Face-centred cubic ground state diagram for negative  $E_1$ . Notation is similar to Fig. 6.10; the two regions marked  $2-\phi$  may be two-phase or polyphase mixtures of  $A_5B$ ,  $A_4B$  and  $A_3B$  superlattices (after Richards and Cahn, 1971).

The ground-state structures for negative  $\xi$  in the more complex f.c.c. diagram (Fig. 6.11) are again two phase, and the  $\text{Cu}_3\text{Au}$  and  $\text{CuAu}$  superlattices are stable. Among the many superlattices predicted for positive  $\xi$  are the known types  $D1a$  ( $\text{Ni}_4\text{Mo}$ ),  $D0_{22}$  ( $\text{Al}_3\text{Ti}$ ) and  $L1_1$  ( $\text{CuPt}$ ), an unknown  $A_5B$  structure, and at  $x \sim 0.5$  an eight-atom cell similar to the  $\text{CuAu II}$  structure (p. 227). Allen and Cahn (1973) have shown that Fig. 6.11 is incorrect in the region  $0 < \zeta < \frac{1}{2}$  and  $\frac{1}{6} < x < \frac{1}{2}$ . The true ground state at  $x = \frac{1}{3}$  is the  $\text{Pt}_2\text{Mo}$  structure (Immm), and for ranges  $\frac{1}{6} - \frac{1}{4}$ ,  $\frac{1}{4} - \frac{1}{3}$  and  $\frac{1}{3} - \frac{1}{2}$  of  $x$  there are 2 or 3-phase, 2-phase and multiphase fields respectively.

For large negative values of  $\xi$  ( $< -\frac{4}{3}$ ) the energy of mixing of the disordered b.c.c. solution is positive even though the ground state is an ordered phase. This also applies to f.c.c. solutions and is contrary to the simple idea that low-temperature ordering is associated with negative  $\Delta_m U$  and phase separation with positive  $\Delta_m U$ . However, the second neighbour interchange energy is unlikely to be larger than the nearest neighbour interchange energy, so the usual association of ordering with the sign of  $\Delta_m U$  is not unreasonable.

Allen and Cahn also investigated the ground state for positive  $\Xi_1$ , and in b.c.c. structures found it to be pure  $A + \text{pure } B$  for all  $\xi > -\frac{2}{3}$  and a mixture of pure  $A$  and the  $B32$  ( $\text{NaTi}$ ) structure for  $\xi < -\frac{2}{3}$ . Similarly for f.c.c. structures,  $A+B$  mixtures are stable for  $\xi > -1$  and mixtures of  $A + \text{the CuPt superlattice}$  for  $\xi < -1$ . Very large second neighbour interactions are thus necessary to produce superlattices when  $\Xi_1$  is positive.

Structures not included in Figs. 6.10 and 6.11 may also be stable because of higher neighbour interactions or non-central forces. Clapp and Moss (1968) assumed that the observed diffuse X-ray maxima due to short-range ordering (see below) indicate minima of the  $k$ -space potential (p. 113), and they deduced the stable structures for stoichiometric compositions with up to third neighbour real-space interactions. Khachatryan (1962, 1973) uses a similar approach based on the symmetry rules of Landau and Lifshitz (1958) for second-order transitions, and especially on the Lifshitz criterion that minima of the harmonic part of the  $k$ -space potential occur at the special points of the Brillouin zone where symmetry elements intersect. He thus considers only structures generated by composition waves with wave vectors corresponding to the special points, and this excludes some f.c.c. superlattices in Fig. 6.11. On the other hand, Khachatryan includes some b.c.c. structures which are not ground states, at least in a near-neighbour model, and which therefore do not appear in the Richards-Allen-Cahn scheme. An elegant discussion of symmetry rules for order-disorder reactions with a comparison of the various approaches is given by de Fontaine (1975).

The use of second neighbour interactions in calculations of partition functions and entropies is very difficult and for the remainder of this chapter we shall use the nearest neighbour model. We now turn to a discussion of the transition from the disordered to the ordered structure, and we begin by distinguishing between the concepts of long-range and short-range order. In most binary superlattices, the tendency of the like atoms to separate and the unlike atoms to attract each other results in all the  $A$  atoms occupying sites comprising one or more sublattices of the whole structure, and the  $B$  atoms occupying sites which make up different sublattices of the structure. The extent to which this ideal arrangement is achieved is a measure of the long-range order of the assembly. Short-range order is a description of the atomic configuration in the immediate vicinity of an atom; if the average

number of unlike neighbours of an  $A$  atom is higher than would be expected for random distribution, the alloy possesses short-range order. Both concepts may be defined in terms of the probability  $P_{AB}(\mathbf{r})$  introduced in Section 23. The alloy possesses short-range order if  $P_{AB}(\mathbf{r})$  is greater than the random value  $2x(1-x)$  for nearest neighbour values of  $\mathbf{r}$ . (Negative short-range order, or clustering, corresponds to  $P_{AB}(\mathbf{r}) < 2x(1-x)$ ; for convenience, any deviation from the random value is often referred to as short-range order, as discussed in the next section.) The alloy possesses long-range order if  $P_{AB}(\mathbf{r})$  for large  $\mathbf{r}$  approaches a limiting value which is not identical with the random value. It is perfectly possible and usual for there to be appreciable short-range order, but no long-range order.

The distinction between long-range and short-range order is sometimes obscured by the possible existence of antiphase domains. These are regions of perfect or nearly perfect order, at the boundaries of which the roles of two or more of the sets of sites are interchanged. Figure 6.12 shows a simple illustration of the effect in a two-dimensional  $AB$  super-

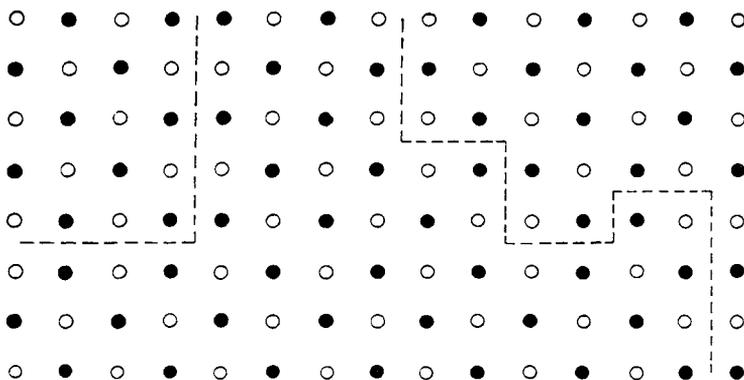


FIG. 6.12. Anti-phase domains in a two-dimensional ordered structure.

lattice. The only atoms not in fully ordered positions are those at the boundaries of the domains, and there is thus a high degree of short-range order. In terms of sublattices which are continuous through all the domains, the long-range order will be small or zero, since there will be nearly equal numbers of the two kinds of atom on each sublattice. Obviously, however, if the domains extend over very many atomic diameters, it would be undesirable to conclude that the structure has no long-range order, and the difficulty can be avoided by considering only a single domain.

Some review articles give a misleading impression that there is no real distinction between a structure of antiphase domains and one possessing short-range order. In fact the domain aggregate may be shown to be thermodynamically unstable, since it involves an increase in energy (at the boundaries) with very little extra entropy to compensate. A structure with genuine short-range order is often the stable configuration because of its high entropy. It is true, of course, that if the size of the antiphase domains continually decreases, the resultant configuration is eventually a state of short-range order only, but at this stage, the domains have lost their individual identity.

A structure of highly ordered antiphase domains is analogous in many ways to a polycrystalline aggregate of strain free grains. The domain boundaries will vanish when true equilibrium is attained, but it is possible in some circumstances to produce a metastable structure in which the domains persist for long periods. From any initial configuration of small domains, the approach to equilibrium occurs in two stages. The domain boundaries first contract, so as to reduce their area, and this is a relatively rapid process, during which the energy decreases continually. It results eventually in an array of planar or nearly planar boundaries, meeting along lines in groups of three at mutual angles of about  $120^\circ$ , and at corners in groups of four. Such an array of either domains or crystal grains is often compared to the structure of a soap froth, and referred to as a foam structure. When it is attained, small displacements of any of the boundaries may lead to an increase in energy.<sup>†</sup> At this stage, the structure changes more slowly, reductions in energy being obtained by movements of boundaries over distances of the order of the domain diameter. As a result of these movements, the smaller domains are progressively eliminated, and the larger domains grow.

There is an important distinction between an array of grains and an array of antiphase domains in a single grain. A grain boundary may separate two crystals of any orientations, but there is a very limited number of different antiphase domains possible with any superlattice structure. Thus in the  $L2_0$  structure, there are only two domains, the corner atoms of the unit cell being occupied by either  $A$  or  $B$  atoms. In an  $A_3B$  superlattice with the  $L1_0$  structure, the  $A$  atoms may occupy any of the four positions of the cubic unit cell, thus giving four different domains.

The foam structure requires the meeting of four different domains at a corner, and hence it is not possible for the  $L2_0$  structure to form such a metastable array. This structure has only one type of antiphase domain boundary, and each such boundary must begin and end on the surface of the grain. The boundaries will straighten rapidly, giving a structure of a few large domains extending right across the crystal. This conclusion that superlattices which have only two types of antiphase domain cannot exist as aggregates of small domains was first pointed out by Bragg (1940).

We have now to formulate our ideas about long- and short-range order in more precise terms, introducing order parameters which are measures of these concepts. In order to evaluate the parameters, we have to use a model, and the nearest neighbour, or Ising, model is the only one which has been found tractable to mathematical analysis. It is useful to note the close analogy between the formation of a phase mixture and the formation of a superlattice, and between the presence of small local aggregates in a solid solution and the presence of short-range order. The analogy is particularly useful for an  $AB$  superlattice, and all results of order-disorder theory in successive approximations may be derived from the corresponding results of the regular solution theory (for a detailed discussion of this point, see Guggenheim, 1952, pp. 113-16). The simplest assumption in the theory of regular solutions does not consider clustering on an atomic scale. The corresponding approxi-

<sup>†</sup> In a fuller discussion, in Section 35, we show that this condition can be satisfied exactly in two dimensions but not in three dimensions. If the structure arises from randomly nucleated centres, the completely metastable array will not be formed, even in two dimensions.

mation in the theory of order-disorder transformations is similarly a theory of long-range order only. This treatment, which we describe first, is usually called the Bragg-Williams theory. These workers originally developed it in a rather different (but equivalent) manner from that which we shall use (Bragg and Williams, 1934, 1935.)

Consider a crystal of a binary alloy having  $N$  sites, of which  $Nx_A$  sites (the  $A$  sites) may be distinguished in some way from the remaining  $Nx_B$  sites (the  $B$  sites). When the alloy is fully ordered, these two sets of sites are completely occupied by  $A$  atoms and  $B$  atoms respectively. The completely disordered state is taken to be the random distribution, in which there will be  $Nx_A^2$   $A$  atoms and  $Nx_Ax_B$   $B$  atoms on the  $A$  sites, and  $Nx_B^2$   $B$  atoms and  $Nx_Ax_B$   $A$  atoms on the  $B$  sites. We define the partially ordered state by the number of "wrong"  $A$  atoms (i.e.  $A$  atoms on  $B$  sites), which we write  $Nx_w$ ; this is also equal to the number of "wrong"  $B$  atoms. The probability that an  $A$  site is occupied by a "wrong" (i.e.  $B$ ) atom is written  $w_A = x_w/x_A$ , and the probability of its being occupied by a "right" atom is  $r_A = 1 - w_A$ . The corresponding probabilities for the  $B$  sites are  $r_B, w_B$ ; note that  $r_A \neq r_B$  except for an equi-atomic alloy. The definition of long-range order introduced by Bragg and Williams, and now generally used, is

$$L = \frac{r_A - x_A}{1 - x_A} = \frac{r_B - x_B}{1 - x_B} = 1 - \frac{x_w}{x_A x_B}. \quad (26.3)$$

This has a maximum value of  $L = 1$  for  $x_w = 0$ , and  $L = 0$  when  $x_w$  has the random value. For  $x_w > x_A x_B$ ,  $L$  becomes negative, but the long-range order must increase again, since this corresponds only to a different labelling of the sites. For example, in an  $L2_0$  superlattice, the situation when all the  $A$  atoms are on the  $B$  sites ( $L = -1$ ) is physically indistinguishable from the fully ordered state of  $L = +1$ . Hence only the magnitude of  $L$  is significant, and we need only consider  $0 \leq x_w \leq x_A x_B$ .

Let us now consider the equi-atomic superlattice in more detail, confining our attention to the simplest condition (type  $L2_0$ ) in which the atomic sites are situated on two equivalent interpenetrating lattices. Each site has  $z$  nearest neighbours, all of which are situated on the other lattice. As there are  $\frac{1}{2}N$  sites of each type,  $w = 2x_w$  and  $L = 2r - 1$ , where the subscripts for  $w, r$  have been dropped, since  $w_A = w_B$ . In the simplest approximation, we assume that the  $Nr/2$   $A$  atoms and the  $Nw/2$   $B$  atoms are distributed completely randomly amongst the  $A$  positions, and similarly for the  $Nw/2$   $A$  atoms and the  $Nr/2$   $B$  atoms on the  $B$  positions. As on p. 188, this enables us to calculate the number,  $zN_{AB}$ , of nearest neighbour pairs of unlike atoms, and thus to obtain an expression for the free energy. A given atom of the first lattice will have on average  $zr$   $B$  neighbours and  $zw$   $A$  neighbours. The numbers of  $A-A$ ,  $B-B$ , and  $A-B$  pairs are thus  $Nzrw/2$ ,  $Nzrw/2$  and  $Nz(r^2 + w^2)/2$ , and the internal energy may be written

$$\begin{aligned} U &= -Nzr(1-r)(\mathcal{E}_{AA} + \mathcal{E}_{BB}) - Nz\{r^2 + (1-r)^2\}\mathcal{E}_{AB} \\ &= -Nz\mathcal{E}_{AB} - Nzr(1-r)\mathcal{E}. \end{aligned} \quad (26.4)$$

We are assuming random distributions of the atoms on each set of sites, and since any arrangement on the  $A$  sites is independent of the arrangement on the  $B$  sites, the number of

distinguishable ways of arranging the atoms is

$$\frac{(N/2)!}{(Nr/2)!(Nw/2)!}$$

After using Stirling's theorem in the usual manner, we obtain for the configurational free energy

$$G \approx F = -Nz\{\mathcal{E}_{AB} + r(1-r)\mathcal{E}\} + NkT\{r \ln r + (1-r) \ln(1-r)\}. \quad (26.5)$$

To find the equilibrium value of  $r$ , we equate  $\partial G/\partial r = 0$ , and obtain

$$(1/(2r-1)) \ln\{r/(1-r)\} = z\mathcal{E}/kT \quad (26.6)$$

or

$$(2/L) \tanh^{-1} L = z\mathcal{E}/kT. \quad (26.7)$$

Equations (26.6) and (26.7) are identical in form with eqns. (23.7) and (23.9), and they have the same properties. For all temperatures, a possible solution is  $r = \frac{1}{2}$  ( $L = 0$ ); this gives a maximum free energy below a critical temperature  $T_\lambda = -z\mathcal{E}/2k$ , and a minimum free energy above this temperature. The degree of long-range order is thus zero above  $T_\lambda$  and increases with falling temperature below  $T_\lambda$ . In the latter range, there are two further roots of eqn. (26.6) symmetrically disposed about  $r = \frac{1}{2}$ . These roots minimize the free energy, and either of them gives the equilibrium value of  $L$  at a given temperature.

In Fig. 6.13 the equilibrium value of  $L$  is plotted as a function of  $T/T_\lambda$ . The curve is nearly

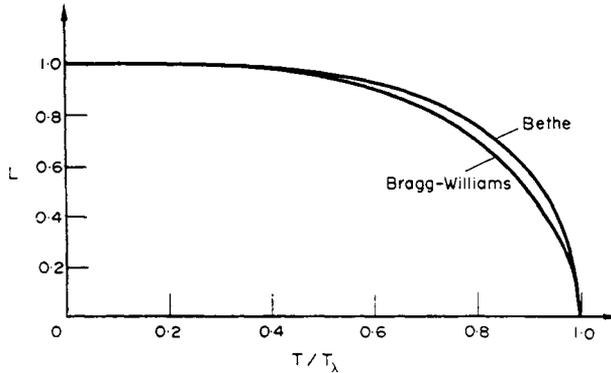


FIG. 6.13. Equilibrium long-range order vs. temperature curve for an  $AB$  superlattice according to the Bragg-Williams and Bethe approximations.

horizontal at first, and then falls more and more steeply as the degree of disorder increases. This behaviour is characteristic of co-operative phenomena, in which the resistance to further disordering (using the term order in a general sense) decreases with decreasing order. The best-known example of such a co-operative phenomenon is the alignment of elementary atomic spins, which leads to ferromagnetism; the Bragg-Williams theory of long-range order is formally analogous to the Weiss theory of ferromagnetism.

The temperature  $T_\lambda$  is often referred to as a  $\lambda$  point in order-disorder and similar theories, the name being taken from the shape of the specific heat anomaly in the neighbourhood of  $T_\lambda$ . From eqn. (26.4), the internal energy may be written

$$U(L) = -Nz\{\bar{\epsilon}_{AB} + (1-L^2)\bar{\epsilon}/4\} \quad (26.8)$$

and the total change in internal energy associated with the ordering process is thus  $-Nz\bar{\epsilon}/4 = NkT_\lambda/2$ . This value is in fair agreement with experiment. The curve of  $U$  as a function of temperature, obtained by combining (26.7) and (26.8), is shown in Fig. 6.14; there is a discontinuity at  $T = T_\lambda$ . The gradient of  $U$  gives the configurational contribution to the specific heat, i.e. the excess specific heat caused by the transformation. This is plotted in Fig. 6.15, together with the experimental curve for the ordering of  $\beta$ -brass, which is an

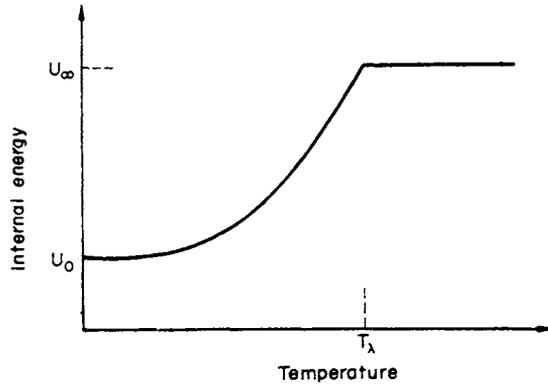


FIG. 6.14. Internal energy of an  $AB$  superlattice according to the Bragg-Williams theory.

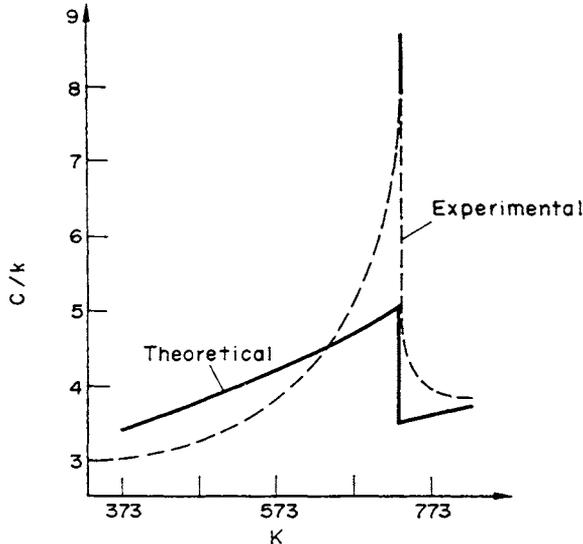
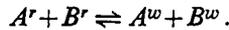


FIG. 6.15. Comparison of the Bragg-Williams prediction of the atomic specific heat with the experimental measurements of Sykes and Wilkinson (1937) on  $\beta$ -brass.

order-disorder transformation of the  $L2_0$ - $A2$  type. In this approximation, the specific heat remains finite at all temperatures, and the ordering reaction is not a first-order phase change.

The original treatment of this theory of long-range order did not refer specifically to the nearest neighbour model, and preceded Bethe's development of this model for alloys. Bragg and Williams considered the change in energy  $W$  associated with the interchange of an  $A$  and a  $B$  atom, both of which were originally in "right" positions, and, finally, in "wrong" positions. Use of Boltzmann's equation then gives an expression for  $r_A/w_A$  in terms of  $W$ ; alternatively this expression may be derived by considering the kinetic balance of the rates at which right atoms interchange to give wrong atoms, and vice versa. These two processes may be written symbolically as an equation (analogous to a chemical equation)



Note that interchanges of right  $A$  atoms with wrong  $B$  atoms, or vice versa, are without significance, since the right or wrong status of neither atom is affected by the interchange.

The rate at which the forward reaction takes place is proportional to the number of right  $A$  atoms,  $N(x_A - x_w)$ , and to the number of right  $B$  atoms,  $N(x_B - x_w)$ , whilst the rate of the back reaction is proportional to  $(Nx_w)^2$ . At equilibrium, these two rates may be equated, and

$$(x_A - x_w)(x_B - x_w)/x_w^2 = k_w/k_r, \quad (26.9)$$

where  $k_w, k_r$  are the rate constants for the forward and back reactions respectively. The equilibrium constant may then be written  $k_w/k_r = \exp(-W/kT)$ .<sup>†</sup> On substituting from eqn. (26.3), the following expression is obtained for the long-range order parameter

$$L = 1 - \frac{[4x_A x_B \{\exp(W/kT) - 1\} + 1]^{1/2} - 1}{2x_A x_B \{\exp(W/kT) - 1\}}. \quad (26.10)$$

The energy  $W$  introduced into this theory is not a constant, but depends upon  $L$ . Clearly  $W$  is zero in a state of complete disorder, since there is then no energy change on interchanging any  $A$  and  $B$  atoms,<sup>‡</sup> and has a maximum value  $W_0$  in the completely ordered state. Bragg and Williams made the simple assumption that there is a linear relation  $W = W_0 L$ . For the equi-atomic superlattice, eqn. (26.10) simplifies to

$$L = 1 - \frac{2 \exp(W/2kT) - 1}{\exp(W/kT) - 1} = \tanh(W/4kT). \quad (26.11)$$

Comparing this with eqn. (26.7), we see that the expressions are identical if  $W = W_0 L = -2z\bar{E}L$ . It may readily be seen that this equivalence applies to the more general expression (26.10) and the corresponding equation derived by the zeroth approximation of the quasi-chemical method. The two theories are in fact identical if the substitution  $W_0 = -2z\bar{E}$  is

<sup>†</sup> Note that this derivation shows the implicit assumption is being made that the entropy of activation is equal for the two processes. For a further discussion, see Part II, Chapter 18.

<sup>‡</sup> More correctly, the energy change is equally likely to be positive or negative.

made, and the Bragg-Williams assumption that  $W = W_0L$  is equivalent to the quasi-chemical assumption of random mixing in the two sets of positions. This serves to emphasise that although the zeroth approximation is developed in terms of the nearest neighbour model, it makes no real use of the short-range character of the binding forces.

We next consider the nature of the changes introduced by removing the restrictive approximation that the atoms are arranged randomly on the two sets of sites. Clearly, it is possible to develop a treatment of order-disorder phenomena based on the assumption of independent nearest neighbour interaction, and this will be exactly analogous to the first approximation of the quasi-chemical method. Such a theory of order-disorder reactions was first given by Bethe (1935). Bethe's analysis was more complex than that given above, but the equivalence of the two methods was proved by Rushbrooke (1938) and Fowler and Guggenheim (1940). We shall not repeat the derivations, but simply assume that we can transcribe the appropriate formulae from Section 24. In particular, we find for the critical temperature of long-range order in an  $L2_0$  superlattice

$$T_\lambda = -\bar{\mathcal{E}}/k \ln(z/(z-2)). \quad (26.12)$$

The equilibrium roots of  $r$  below  $T = T_\lambda$  are given by

$$\frac{1 - [r/(1-r)]}{[r/(1-r)]^{1/z} - [r/(1-r)]^{(z-1)/z}} = \exp(-\bar{\mathcal{E}}/kT), \quad (26.13)$$

which is the equivalent of (24.16). In Fig. 6.13 we compare the curve for  $r$ , and hence for  $L = 2r - 1$ , given by eqn. (26.13) for an  $L2_0$  superlattice with that given by eqn. (26.6). Both equations predict that the long-range order falls to zero at a critical temperature ( $\lambda$  point), but the Bethe theory shows that the long-range order decreases more slowly than is predicted by the Bragg-Williams approximation. It follows from the general correspondence of the methods, that the order-disorder curves of Fig. 6.13 are the same as the solubility limit curves of Fig. 6.9.

The configurational internal energy will be given by

$$U = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - 2\bar{N}_{AB}\bar{\mathcal{E}}/N) \quad (26.14)$$

(see p. 191), where  $\bar{N}_{AB}$  is a function of temperature and is determined by eqn. (24.7) with  $N_A = N_B = \frac{1}{2}N$ . As  $T \rightarrow 0$  K, the superlattice structure becomes perfect, and  $\bar{N}_{AB} \rightarrow \frac{1}{2}N$ . Thus at 0 K

$$U_0 = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - \bar{\mathcal{E}}). \quad (26.15)$$

At very high temperatures,  $\bar{N}_{AB}$  is equal to the value  $N_{AB}^\theta$  corresponding to  $r = \frac{1}{2}$  and random mixing on each of the sublattices, i.e. to  $N/4$ . We then have

$$U_\infty = -\frac{1}{2}Nz(\bar{\mathcal{E}}_{AA} + \bar{\mathcal{E}}_{BB} - \frac{1}{2}\bar{\mathcal{E}}), \quad (26.16)$$

and the total energy of disordering is  $-Nz\bar{\mathcal{E}}/4$ , as in the zeroth approximation. The present treatment shows, however, that only part of the configurational energy of ordering disappears below the  $\lambda$  point. From eqns. (26.14), (26.15), and (26.16)

$$(U_T - U_0)/(U_\infty - U_0) = 2(1 - 2\bar{N}_{AB}/N). \quad (26.17)$$

This equation gives implicitly the value of the configurational energy  $U_T$  as a function of  $T$  or of the order parameter  $L$ . At any temperature,  $\overline{N_{AB}}$  may be found from the following equation, which is analogous to eqn. (24.7),

$$\xi^2 = (r - \xi)(1 - r - \xi) \exp(2\mathcal{E}/kT),$$

where  $2\xi = 1 - 2\overline{N_{AB}}/N$ . The solution may be written in the form (24.8) as

$$\xi = 2\xi/(\beta + 1) = 2r(1 - r)/(\beta + 1),$$

and  $\beta$  is given by eqn. (24.9) with  $\mathcal{E}$  replaced by  $-\mathcal{E}$  and  $x$  replaced by  $r$ . The value of  $r$  in these equations is given by (26.6). At  $T = T_\lambda$  and higher temperatures,  $r = \frac{1}{2}$  and  $\beta = \exp(-\mathcal{E}/kT)$ . This gives

$$\overline{N_{AB}} = \frac{N \exp(-\mathcal{E}/kT)}{2[1 + \exp(-\mathcal{E}/kT)]}. \quad (26.18)$$

Above the critical point, the curve of  $U_T$  against  $T$  is represented by the equation

$$U_T - U_0 = \frac{-Nz\mathcal{E}}{2[1 + \exp(-\mathcal{E}/kT)]}. \quad (26.19)$$

Below the critical point, it is not possible to write a simple analytical expression for  $U_T - U_0$ . The curve of  $U_T - U_0$  is plotted against  $T/T_\lambda$  in Fig. 6.16, together with the corres-

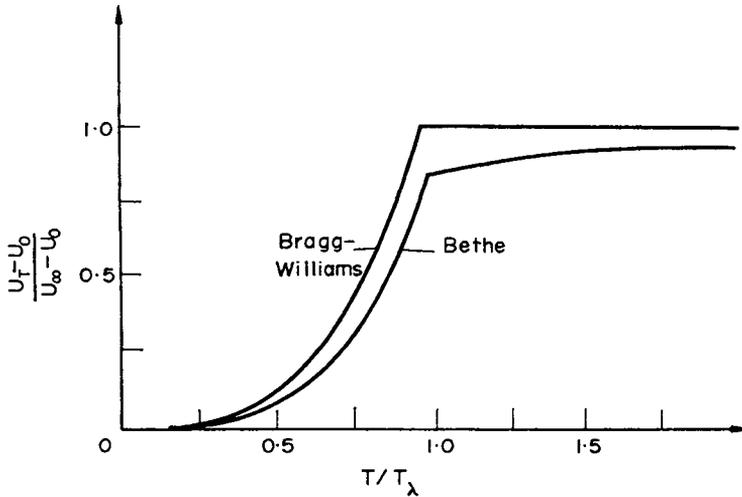


FIG. 6.16. Comparison of the internal energy vs. temperature curves given by the zeroth (Bragg-Williams) and first (Bethe) approximations for an  $AB$  superlattice.

ponding curve for the zeroth approximation. We may also calculate from (26.19) that

$$(U_\lambda - U_0)/(U_\infty - U_0) = (z - 2)/(z - 1).$$

For a b.c.c. structure, sixth-sevenths of the total ordering energy is destroyed below the critical temperature according to this approximation, and the remaining one-seventh dis-

appears more gradually as the temperature is raised above  $T_\lambda$ . The gradient of  $U_T$  gives the configurational specific heat, i.e. the excess specific heat above the Debye curve. This is shown in Fig. 6.17.

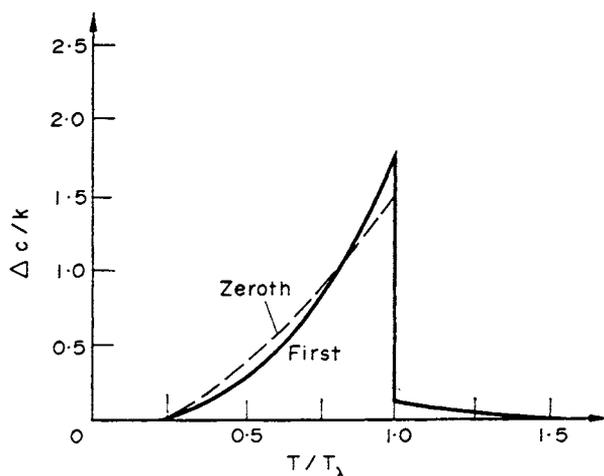


FIG. 6.17. Comparison of the excess atomic specific heat curves given by the zeroth and first approximations.

The principal advantage of the Bethe treatment over the Bragg-Williams approximation is in this prediction of an excess specific heat above the disordering temperature, since this is always observed experimentally. The physical interpretation of the effect is in terms of the concept of short-range order. Although the long-range order becomes zero at the critical temperature, the number of unlike nearest neighbour pairs remains larger than would be found in a random distribution, so that short-range order persists above the critical point. The destruction of long-range order has been frequently compared to the melting of a solid; local crystallinity persists in the liquid, and local order in the disordered solution.

Bethe introduced a short-range order parameter  $\sigma$ , which is defined so that  $\sigma = 1$  for perfect long-range order, and  $\sigma = 0$  for a completely random distribution. In these respects,  $\sigma$  is equivalent to  $L$ , but the definition is directly in terms of the nearest neighbours of an atom rather than of the segregation into different lattices. The general definition of the short-range order parameter is

$$\sigma = (\overline{N_{AB}} - N_{AB}^0) / (N_{AB}^0 - N_{AB}^1), \quad (26.20)$$

where  $\overline{N_{AB}}$ ,  $N_{AB}^0$  are the actual value of  $N_{AB}$  and the value for a purely random arrangement, as before, and  $N_{AB}^1$  is the value of  $N_{AB}$  for the fully ordered state. In this simple case we have been considering,  $N_{AB}^0 = N/2$ , and  $N_{AB}^1 = N/4$ , so that

$$\sigma = (4\overline{N_{AB}}/N) - 1. \quad (26.21)$$

Comparing this with eqn. (26.17), we see

$$(U_T - U_0)/(U_\infty - U_0) = 1 - \sigma. \quad (26.22)$$

With this definition of ordering, the internal energy of ordering is proportional to the degree of short-range order. At  $T_\lambda$ , for example, the short-range order parameter  $\sigma = \frac{1}{7}$ .

Although Bethe's theory is successful in explaining short-range order, detailed comparison of experimental and theoretical results shows rather poor agreement. Attempts at more exact theories usually take the form of higher cluster approximations or series expansions (see Takagi, 1941). We shall not consider these here, but it is interesting to refer briefly again to the exact solutions obtained in two dimensions. A square lattice forms a superlattice of the type described above, there being no  $AA$  or  $BB$  pairs in the fully ordered state. The rather surprising result, first obtained by Onsager (1944) and discussed more fully by Wannier (1945), is that the specific heat of the structure has a logarithmic infinity at  $T = T_\lambda$ . The heat content remains continuous, i.e. there is no latent heat, and the change is thermodynamically of the second order (see pp. 226-8). This result is at variance with the predictions of all the approximate calculations, according to which there is a finite discontinuity in the specific heat. Moreover, in the exact two-dimensional solution, the curve is symmetrical about  $T = T_\lambda$ , so that as much of the energy of disordering is required above  $T_\lambda$  as below it. Exact solutions in three dimensions are unlikely to be obtained; it is probable that such a solution would still contain a logarithmic infinity in the specific heat, but would not be symmetrical about  $T_\lambda$ .

The above discussion has been confined to the simplest equi-atomic superlattice. The next most important example is the structure of type  $L1_2$ , formed in  $A_3B$  alloys. We shall not develop the theory of this superlattice, but shall merely describe the results. We now have  $w_B = 3w_A = 4x_w$ , and the Bragg-Williams definition of long-range order becomes

$$L = (4r_A - 1)/3 = 1 - 16x_w/3. \quad (26.23)$$

Attempts to apply the first approximation, or Bethe theory, to the  $L1_2$  superlattice result in a contradiction; the hypothesis of the non-interference of nearest neighbour pairs does not lead to equilibrium long-range order at all. The reason for this difficulty seems to be that at least four atom sites are needed to define a unit cell of the superlattice. It is thus necessary to use a higher approximation of the quasi-chemical theory (cluster variation method), embodying the hypothesis of the non-interference of tetrahedral groups of atoms. The stability of the  $L1_2$  structure at the  $A_3B$  composition was first shown in this way by Yang (1945), and the treatment was considerably extended in later papers by Yang and Li (1947) and Li (1949). These papers also considered the more difficult problems associated with non-stoichiometric compositions. When  $x_B$  is a variable, the use of the quasi-chemical method ensures that all properties are symmetrical about  $x_B = \frac{1}{2}$ , so, in this treatment, superlattices based on both  $A_3B$  and  $AB_3$  necessarily appear in the same binary alloys.

The results of the Bragg-Williams approximation and of the tetrahedral cluster method are shown in Fig. 6.18, which gives the equilibrium degree of long-range order as a function of temperature. In contrast to the  $L2_0$  type of superlattice,  $L$  drops discontinuously to zero at the critical temperature. The transformation thus requires a latent heat, and is correctly

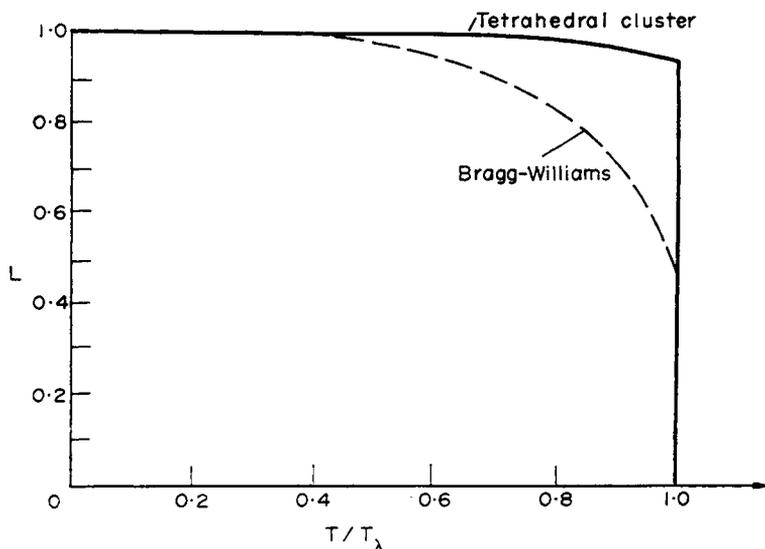


FIG. 6.18. Equilibrium long-range order vs. temperature curve for an  $A_3B$  superlattice of type  $L1_2$  according to the Bragg-Williams and tetrahedral cluster approximations (after Guggenheim, 1952).

described as a thermodynamic phase change of the first order. The transformation temperature, at which the superlattice phase with a finite degree of long-range order is in equilibrium with the solution of zero long-range order, is given by

$$\left. \begin{aligned} T &= -z\bar{E}/7.3k && \text{(Bragg-Williams),} \\ T &= -z\bar{E}/14.6k && \text{(tetrahedral cluster).} \end{aligned} \right\} \quad (26.24)$$

These values are obtained by numerical solution of the equations, no simple analytical expressions being obtained.

An approximate theory of order in alloys due to Cowley (1950 a, b) differs in some respects from the cluster variation methods. Cowley considers a set of short-range order coefficients which effectively define the probability  $P_{AB}(\mathbf{r})$  for each of the different shells (successively larger values of  $r$ ) around an atom. The internal energy is defined in terms of the interactions of pairs of atoms in different shells, so that the theory is not confined to nearest neighbour interactions. It is assumed as an approximation that the order coefficients are independent of each other, this corresponding approximately to the independence of pairs in the first approximation of the quasi-chemical theory.

Cowley's theory has the advantage of being related rather more closely to the experimental X-ray measurements of the  $P_{AB}(\mathbf{r})$  quantities than are the above descriptions. The results of work on the diffuse X-ray scattering from solid solutions are frequently expressed in terms of the order coefficients used by Cowley (positive or negative for short-range order and clustering; see Section 27), together with size coefficients representing the displacements of the atoms from the ideal sites of a structure with the measured lattice parameter. This work will not be described here; reviews have been given by Averbach (1956) and

Sivertsen and Nicholson (1961). From the mean pair probabilities for  $\text{Cu}_3\text{Au}$ , Cowley deduced that the values of  $\bar{E}$  for the first three shells (sets of nearest neighbours) are given by  $\bar{E}_1/k = 358^\circ$ ,  $\bar{E}_2/k = -34^\circ$ ,  $\bar{E}_3/k = -19^\circ$ . For the  $L2_0$  superlattice, Cowley's theory gives the same results as the Bragg-Williams theory; the disordering temperature for the other types of superlattice differs slightly from the predictions of the above theories.

The question of whether superlattice changes are first- or second-order thermodynamic transformations has attracted much discussion. The approximate solutions of the nearest neighbour model suggest that some transitions involve a continuous decrease of the long-range order parameter to zero (second-order change), whilst others predict a discontinuous fall at  $T = T_c$  (first-order change). Thus changes from  $L2_0$  or  $D0_3$  structures to disordered b.c.c. structures may be second order, whilst most others are first order. In these second-order changes, both superlattice and disordered phases are non-close-packed structures, in which the nearest neighbours of any one atom are not nearest neighbours of each other. In his review article on order-disorder phenomena in metals, Guttman (1956) considerably clarified these difficult problems, and the following description is taken mainly from his original work.

The thermodynamic classification of transitions is based on the order of the lowest derivative of  $G$  which shows a discontinuity at the transition temperature. If the values of this function are  $G^\alpha$  and  $G^\beta$  for the two forms concerned, the transition temperature is defined by the condition  $G^\alpha = G^\beta$ . In a normal first-order transition, there are discontinuities in the derivatives of  $G$  with respect to temperature (entropy) and with respect to pressure (volume). In a second-order transition, the entropy and volume are continuous, but their derivatives (e.g. the specific heat and the compressibility) are not. It follows that the enthalpy is also continuous, and the transition has an anomalous specific heat but no latent heat.

A well-known difficulty in the theory of even-order transitions arises if the difference in the free energies of the two phases  $\Delta G^{\alpha\beta}$  is expanded as a Taylor series in terms of the difference in temperature  $\Delta T = T - T^{\alpha\beta}$ . This gives

$$\Delta G^{\alpha\beta} = -(\Delta S^{\alpha\beta}) \Delta T + \frac{1}{2}(\partial^2 G / \partial T^2) \Delta T^2 + \frac{1}{6}(\partial^3 G / \partial T^3) \Delta T^3 + \dots$$

The sign of  $\Delta G^{\alpha\beta}$  is determined by the sign of the first non-vanishing term on the right, and this is the  $n$ th term for an  $n$ th-order transition. For a first- or third-order transition,  $\Delta G^{\alpha\beta}$  changes sign with  $\Delta T$ , but for an even-order transition, it has the same sign above and below  $T^{\alpha\beta}$ . Thus if attempts are made to draw  $G^\alpha$  and  $G^\beta$  curves, they do not intersect at the transition temperature, but only touch, with one curve always below the other. This apparently results in either  $\alpha$  or  $\beta$  always being stable.

This difficulty is thought to arise because of the attempt to extrapolate properties of the assembly into regions in which they do not exist. We cannot really draw separate  $G^\alpha$  and  $G^\beta$  curves for a second-order superlattice transition, but only a single curve which represents an ordered structure below the transition temperature and a substantially disordered structure above it. The existence of the singularity in  $\partial^2 G / \partial T^2$  renders the analytic expansion above invalid.

A second-order transition in a solid solution implies not only a vanishing of the latent

heat of transition, but also a continuity of composition. The transition temperature is marked by a single line on the phase diagram, and the coexistence of ordered and disordered phases is not possible at equilibrium. This provides one of the most convenient experimental tests of the order of a transition, it being very difficult in practice to distinguish between latent heats and anomalous specific heats. In the early stages of order-disorder theory, there was a tendency to regard the disordering of almost all superlattices as second-order transitions, but it has become increasingly evident that many of these changes are first order. It seems obvious that the transition must be first order when there is a change in lattice symmetry, as in the  $L1_0$  (tetragonal) superlattice formed from the f.c.c. structure, and the most convincing evidence of the existence of equilibrium two phase regions comes from alloys undergoing this structural change.†

Rhines and Newkirk (1953) have suggested that all superlattice transitions are first-order phase changes, and they presented some evidence, based on electrical resistivity measurements, to show that this is true even for the  $L2_0$ - $A2$  transition in copper-zinc alloys. In general, this suggestion has not been substantiated by later workers; for example, a two-phase region was sought but not found in a very careful investigation of copper-zinc by Beck and Smith (1952). Moreover, measurements of long-range order by X-ray methods tend to support the division of superlattice changes into first- and second-order types. For the  $L2_0$  structures in copper-zinc and silver-zinc alloys, the long-range order decreases continually with temperature (Chipman and Warren, 1950; Muldower, 1951), the decrease becoming more and more rapid as the critical temperature is approached. Whilst the experimental techniques do not allow the definite conclusion that there is no discontinuity in  $L$ , it is certain that the magnitude of any discontinuous drop is much smaller than that for the transitions from  $L1_0$  or  $L1_2$  superlattices in copper-gold and copper-platinum alloys. In these alloys, the amount of long-range order decreases only slightly as the temperature is increased within the superlattice region, and then drops discontinuously to zero (Keating and Warren, 1951; Roberts, 1954; Walker, 1952). Thus these measurements tend to confirm the qualitative predictions of the nearest neighbour model for all kinds of superlattice.

We have assumed throughout this section that the ordering tendency is provided by the chemical interactions between nearest neighbour atoms. Suggestions are often made that an atomic size disparity is an important factor in lowering the energy of the superlattice relative to that of a random arrangement, a view first proposed by Hume-Rothery and Powell (1935). It is not entirely clear whether this distinction is meaningful (see p. 201), but it is obvious that a strain energy calculation of the type used in Section 25 has no relevance to the ordering energy, which must arise from more localized interactions. At the same time, there are a number of superlattices known to have large unit cells, and the formation of these structures cannot be explained on the basis of purely nearest neighbour interactions.

The best-known example of a superlattice with a large unit cell is the structure CuAu II. At low temperatures, equi-atomic copper-gold alloys form a tetragonal superlattice of the  $L1_0$  form (CuAu I) and, at high temperatures, the structure is disordered f.c.c. However,

† The Bragg-Williams theory predicts a second-order transition, but the first approximation of the quasi-chemical theory correctly predicts a first-order transition (Guggenheim, 1952).

it was found in 1936 that in an intermediate temperature range, now known to be  $\sim 380^\circ\text{C}$  to  $\sim 405^\circ\text{C}$ , the equilibrium structure has a more complex unit cell with dimensions  $10a$ ,  $a$ , and  $c$ , where  $c$  and  $a$  are the tetragonal parameters (Johansson and Linde, 1936). The structure may be considered to be formed by introducing antiphase domain boundaries parallel to (100) planes at every five lattice planes of the tetragonal structure, so that the repeat unit has to contain two opposite "domains". The complete diffraction effects, which are rather complex, also indicate a small expansion across the antiphase domain boundaries. Direct evidence for the existence of these domains has been obtained by transmission electron-microscopy of thin films (Pashley and Presland 1958–9; Glossop and Pashley, 1959). The structure consists of regions within which the antiphase domain boundaries are parallel planes; two kinds of region are observed, corresponding to the two perpendicular  $a$  directions of the tetragonal cell.

Other examples of structures with large unit cells are now known, and are formed by a similar disturbance of the structure of a simple superlattice, but the spacing of the antiphase boundaries is not always constant, sometimes varying with temperature and composition (Schubert *et al.*, 1954, 1955). From this point of view, it seems attractive to regard the structures as superlattices containing some kind of fault. Nevertheless, and in spite of our description of the CuAu II structure as a tetragonal superlattice with regularly spaced antiphase domain boundaries, it is quite clear that, in these alloys at least, the large superlattice is the thermodynamically stable form. This cannot be explained by the nearest neighbour model because an  $A$  atom has the same number of nearest neighbour  $B$  atoms in both the CuAu I and the CuAu II structures if the slight deviations from cubic symmetry are ignored.

The thermodynamics of the transitions CuAu I  $\rightarrow$  CuAu II  $\rightarrow$  disordered f.c.c. have been investigated by Oriani and Murphy (1958). The latent heats for the two changes are 888 and 1590 J g atom $^{-1}$  respectively; the former value seems remarkably large in view of the slight structural rearrangement involved. The results show other disagreements with the quasi-chemical theory. For example, the total disordering energy per g-atom for the transition CuAu I  $\rightarrow$  disordered f.c.c. can be evaluated, and the temperature  $T_1$  for this transition in the absence of the CuAu II structure can be estimated. The ratio of the disordering energy to  $RT_1$  is found to be 0.73 compared with 0.5 of the zeroth approximation and 1.37 given by Li's treatment of the  $L1_2-A1$  change using the tetrahedral cluster method. The heats of formation of both ordered and disordered phases were measured by Oriani and Murphy, and give independent estimates of  $\bar{E}$ . These values are 1570 and 1940 J g atom $^{-1}$  respectively. It thus seems clear that the quasi-chemical treatment is extremely unsatisfactory for the CuAu superlattices.

## 27. FLUCTUATIONS IN SOLID SOLUTIONS: SHORT-RANGE ORDER AND CLUSTERING

We have emphasized that the pair probability functions will generally not have the values corresponding to a completely random arrangement, even in a substantially disordered solid solution. These functions, or related order parameters, may be measured by X-ray methods, and give quantitative information on the amount of short-range order or

clustering. In general, short-range order, for which  $P_{AB}(\mathbf{r}) > 2x(1-x)$  when  $\mathbf{r}$  represents a nearest neighbour vector, is expected if the solid solution undergoes a superlattice transition at lower temperatures. Clustering (negative short-range order), for which  $P_{AB}(\mathbf{r}) < 2x(1-x)$  for nearest neighbour vectors, is expected if there is a solubility gap at lower temperatures.<sup>†</sup>

The X-ray measurements of short-range order give the average number of neighbours of given kinds possessed by any atom. We now examine the nature of the qualification expressed by the word "average". An instantaneous picture of a solid solution would obviously show that many atoms had a surrounding configuration quite different from the average. This is because the average configuration is obtained by a dynamic averaging process over all the available configurations. In postulating this, we are assuming some mechanism for changing from one configuration to another; this is achieved in practice by atomic diffusion.

For kinetic applications, we require a quantitative estimate of how the arrangement of a small region of the assembly may momentarily differ from the average arrangement. This is provided by the theory of fluctuations, and we express the probable state of the assembly in terms of the root mean square deviation from the average arrangement. We begin by considering a solid solution in which the atomic arrangement is completely random. The statistical properties of such a solution are especially easy to derive, since all arrangements have equal probability.

We consider in all  $N$  atoms, each having  $z$  nearest neighbours. For each atom, we fix attention first on a particular neighbour. The probability that this neighbour is a  $B$  atom is  $x$ , and hence there will be  $Nx$  of the neighbours we are considering which are  $B$  atoms. Now consider another neighbour of each of our  $N$  atoms. Since the probabilities are independent, there will again be a chance  $x$  that each such neighbour is a  $B$  atom. The number of  $B-B$  pairs among the  $N$  sets of two neighbours will thus be proportional to  $Nx^2$ , whereas the number of  $A-B$  pairs will be  $Nx(1-x)$ . Similarly, the number of  $A-A$  pairs and of  $A-B$  pairs will be proportional to  $N(1-x)^2$  and  $Nx(1-x)$ . We thus have that for  $z = 2$ , the numbers of atoms having two  $A$  atoms, an  $A$  atom and a  $B$  atom, and two  $B$  atoms as nearest neighbours are  $N(1-x)^2$ ,  $2Nx(1-x)$  and  $Nx^2$  respectively. Exactly similar reasoning shows that for  $z = 3$ , the numbers of atoms with three  $A$  atoms, two  $A$  atoms and one  $B$  atom, one  $A$  atom and two  $B$  atoms, and three  $B$  atoms are proportional to  $N(1-x)^3$ ,  $3N(1-x)^2x$ ,  $3N(1-x)x^2$  and  $Nx^3$  respectively. The relative numbers of neighbours of the different types are given by the terms of the expansions  $\{(1-x)+x\}^2$  and  $\{(1-x)+x\}^3$ . We conclude by extension of this reasoning that the relative numbers of atoms which have as nearest neighbours  $zA$  atoms,  $(z-1)A$  atoms and one  $B$  atom,  $(z-2)A$  atoms and two  $B$  atoms, . . . , one  $A$  atom and  $(z-1)B$  atoms, and  $zB$  atoms are given by the terms of the binomial expansion

$$\{(1-x)+x\}^z.$$

<sup>†</sup> For many years, gold-nickel alloys were quoted as an extreme example of the failure of the quasi-chemical theory, since they have a solubility gap at low temperatures, and X-ray measurements (Flinn, *et al.*, 1953) indicated positive short-range order at high temperatures. Later work by Munster and Sagel (1959) has shown that the interpretation of the earlier X-ray results may have been incorrect, and there may actually be clustering above the solubility gap. It is possible that this is a necessary result, and is independent of the assumptions of any particular model.

This expression is not a continuous mathematical function, i.e. there is no chance of finding an atom with a fractional number of  $A$  atoms among its nearest neighbours. There is a distinction between the average number of neighbours of a given type and the most probable number. Thus if  $x = 0.7$  and  $z = 4$ , each atom has on the average 1.2  $A$  atoms and 2.8  $B$  atoms as nearest neighbours. The most probable neighbours for any atom are one  $A$  atom and three  $B$  atoms.

The above result is not confined to nearest neighbour probabilities, since we have simply calculated the probability of having a given number of  $B$  atoms amongst the  $z$  atoms which happen to be neighbours of any atom. We are often interested in rather larger clusters of (say)  $n$  atoms. The probabilities of our finding 0, 1, 2, 3, ...,  $n-1$ ,  $n$   $B$  atoms in such a cluster when the average composition is  $x$  are given by the terms of the expansion  $\{(1-x)+x\}^n$ , and the general term, i.e. the probability of the group containing (say)  $m$   $B$  atoms is

$$\binom{n}{m} (1-x)^{n-m} x^m. \quad (27.1)$$

This formula may be used as a first approximation in the problem of nucleus formation in a solid solution.

If the solution is very dilute, the above law may be simplified. We let  $x \rightarrow 0$ , increasing  $n$  at the same time, so that the average number of solute atoms in the group,  $nx$ , remains constant. The general term of the binomial expansion can now be written as

$$\frac{n!}{m!(n-m)!} \left(\frac{nx}{n}\right)^m \left(1 - \frac{nx}{n}\right)^{n-m} = \frac{(nx)^m}{m!} \left(1 - \frac{nx}{n}\right)^n \left[ \frac{n!}{(n-m)! n^m (1 - nx/n)^m} \right].$$

By using Stirling's theorem, the term in square brackets reduces to

$$\left(1 - \frac{nx}{n}\right)^{-m} \left(1 - \frac{m}{n}\right)^{-n+m-\frac{1}{2}} \exp(-m)$$

and as  $n \rightarrow \infty$  with  $nx$  constant, this may be replaced by unity. At the same time, the term  $(1 - nx/n)^n$  tends to  $\exp(-nx)$ . We thus have, finally, for the probability of finding  $mB$  atoms amongst  $n$  atoms:

$$\{(nx)^m/m!\} \exp(-nx). \quad (27.2)$$

This is Poisson's distribution. The expression has a maximum value when  $m = nx$ , i.e. when the number of solute atoms in the group equals the average number, and the concentration fluctuation is zero. The fluctuation is most conveniently expressed by the root mean square deviation from the average value; this is the conventional mean deviation of statistical theory.

Since  $(m - nx)^2 = m^2 - (nx)^2 + 2nx(nx - m)$ , and the mean deviation of  $(nx - m)$  is zero, we have

$$\overline{(m - nx)^2} = \sum_{m=0}^n [m^2 - (nx)^2] \{(nx)^m/m!\} \exp(-nx).$$

The first part of this expression is

$$\sum_{m=0}^n (m^2/m!) (nx)^m \exp(-nx) = nx(1+nx),$$

so that  $\overline{(m-nx)^2} = nx$ , and the mean deviation is  $(nx)^{1/2}$ . Expressed as a composition fluctuation, we write  $m = n(x + \Delta x)$ , and the mean deviation is given by

$$\overline{(\Delta x)^2} = x/n. \quad (27.3)$$

The concentration fluctuations thus increase as the number of atoms in the group diminishes. Provided  $n$  is large enough (or, equivalently,  $x$  is small enough), to justify the use of Poisson's equation, the mean deviation is proportional to  $n^{-1/2}$ .

A similar method may be used to find the mean deviation of the more exact binomial expansion. We then obtain

$$\text{or } \left. \begin{aligned} \overline{(m-nx)^2} &= nx(1-x) \\ \overline{(\Delta x)^2} &= x(1-x)/n. \end{aligned} \right\} \quad (27.4)$$

Obviously, this reduces to (27.3) when  $x$  is small.

In the above discussion, we have not been concerned with the effects of temperature at all. This is because all arrangements of the ideal solution have the same energy. Temperature has no effect on the magnitude of the fluctuations, although it will determine the rate of fluctuation. In a real solution, we must consider the change in free energy which is associated with a fluctuation. The probability of the fluctuation is then related to this change in energy by the Boltzmann factor.

Consider a region of solid solution which has initially a concentration of  $B$  atoms equal to the average or equilibrium concentration,  $x$ . As a result of a fluctuation, one half of this region may be supposed to contain a fraction  $x + \Delta x$  of  $B$  atoms, whilst the other half contains a fraction  $x - \Delta x$  of  $B$  atoms. If the free energy per atom is  $g(x)$ , we have

$$\Delta g = \frac{1}{2} \{g(x + \Delta x) + g(x - \Delta x)\} - g(x).$$

And after expanding in a Taylor series, we obtain

$$\Delta g = \frac{1}{2} (\partial^2 g / \partial x^2) (\Delta x)^2$$

provided that  $\Delta x$  is small. If the original volume element contained  $n$  atoms, the total change in the free energy of the assembly is  $n \Delta g$ . Since the average value of  $n \Delta g$  must be  $\frac{1}{2} kT$ , the mean square fluctuation is

$$\overline{(\Delta x)^2} = kT/n(\partial^2 g / \partial x^2) \quad (27.5)$$

and this equation replaces (27.4) when the energy of the assembly is a function of the arrangement. The value of  $\partial^2 g / \partial x^2$  for the zeroth approximation of the quasi-chemical theory is derived from eqn. (23.5). Substituting this into (27.5),

$$\overline{(\Delta x)^2} = \left( \frac{n}{x(1-x)} - \frac{2z\bar{\epsilon}}{kT} n \right)^{-1}. \quad (27.6)$$

This equation reduces to (27.4) when  $\mathcal{E} \rightarrow 0$ , this being the condition for the solution to be ideal. A similar result is obtained when  $T$  becomes very large, so that, at high temperatures, the fluctuations approach those expected in a random solid solution, as is physically obvious. At lower temperatures, the fluctuations depend on the sign of  $\mathcal{E}$ . If  $\mathcal{E}$  is positive, there is a tendency to clustering, and the fluctuations are greater than those in an ideal solution. As the temperature is lowered, the fluctuations become larger, and according to (27.5) are infinite when  $\partial^2 g / \partial x^2 = 0$ . This equation is, of course, only valid for small  $x$ , but it is clear that large fluctuations may be expected under these conditions since there is no first-order dependence of the free energy on the atomic arrangement. Thus when the spinodal is reached, all concentrations have the same probability. When  $\mathcal{E}$  is negative (tendency to ordering) the fluctuations are always smaller than would be found in an ideal solution.

The above treatment, though giving a better approximation to the fluctuations in a solution than do the purely statistical expressions, is logically rather unsatisfactory. We have used the zeroth approximation for the free energy, and so have assumed that the equilibrium configuration is completely random, even though the energy depends on this configuration. In a more accurate treatment,  $\partial^2 g / \partial x^2$  would be calculated from the first approximation, and used in eqn. (27.5). Unfortunately, this does not lead to any simple analytical expression. We should note, however, the difference between the value of  $(\overline{\Delta x})^2$  given by eqn. (27.6) and the mean number of neighbours of a given kind possessed by any atom. Thus above the critical temperature, the equilibrium, or mean, composition in any region of  $n$  atoms

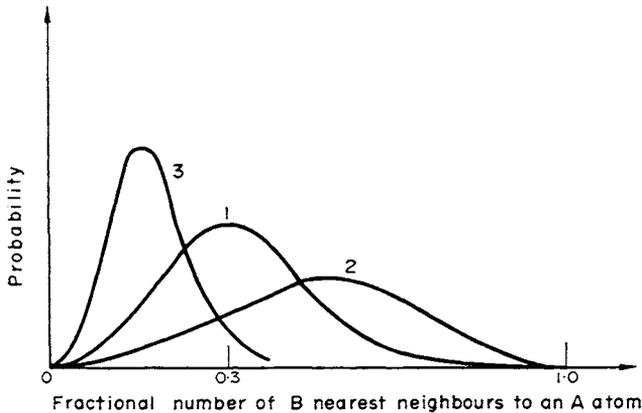


FIG. 6.19. Probability that a given fraction of the nearest neighbours of an  $A$  atom are  $B$  atoms in a solid solution for which  $x = 0.3$  (after Smoluchowski, 1951). Curve 1: Ideal solution, Curve 2: Short-range order. Curve 3: Clustering.

is always  $x$  so long as  $n$  is large. For smaller values of  $n$ , the concentration fluctuations become correspondingly large, and (27.4) ceases to apply. We have already seen that for  $\mathcal{E}$  negative, the tendency to order persists above the critical temperature, and the number of  $A-B$  pairs is greater than the random value. Exactly the same reasoning holds for  $\mathcal{E}$  positive; we expect a finite configurational specific heat corresponding to the atomic clusters which gradually disappear with increasing temperature. This clustering or short-range

ordering is quite distinct from the fluctuations described above, and expresses the fact that if we view on a sufficiently fine scale, the average distribution is no longer uniform. For  $\bar{\epsilon}$  positive we have a short-range phase mixture, just as we have short-range order for  $\bar{\epsilon}$  negative. Thermodynamically, the solution remains one phase, since thermodynamic concepts only have meaning for large numbers of atoms. Figure 6.19 shows qualitatively the number of nearest neighbours of type  $A-B$  for  $x = 0.3$ .

## REFERENCES

- ALLEN, S. M. and CAHN, J. W. (1972) *Acta Metall.* **20**, 423; (1973) *Scripta Metall.* **7**, 1261.  
 AVERBACH, B. L. (1956) *Theory of Alloy Phases*, p. 301, American Society for Metals, Cleveland, Ohio.  
 BECK, L. H. and SMITH, C. S. (1952) *Trans. Am. Inst. Min (Metall.) Engrs.* **194**, 1079.  
 BETHE, H. (1935) *Proc. R. Soc. A*, **150**, 552.  
 BRAGG, W. L. (1940) *Proc. Phys. Soc.* **52**, 105.  
 BRAGG, W. L. and WILLIAMS, E. J. (1934) *Proc. R. Soc. A*, **145**, 699; (1935) *Ibid.* **151**, 540.  
 BROWN, W. B. and LONGUET-HIGGINS, H. C. (1951) *Proc. R. Soc. A*, **209**, 416.  
 CAHN, J. W. (1961) *Acta Metall.* **9**, 795; (1969) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 1, Institute of Metals, London.  
 CAHN, J. W. and HILLIARD, J. E. (1958) *J. Chem. Phys.* **28**, 258; (1971) *Acta Metall.* **19**, 151.  
 CHIPMAN, W. B. and WARREN, B. E. (1950) *J. Appl. Phys.* **21**, 696.  
 CLAPP, P. C. and MOSS, S. C. (1968) *Phys. Rev.* **171**, 754.  
 COTTRELL, A. H. (1955) *Theoretical Structural Metallurgy*, Arnold, London.  
 COWLEY, J. M. (1950a) *J. Appl. Phys.* **21**, 24; (1950b) *Phys. Rev.* **77**, 664.  
 DE FONTAINE, D. (1975) *Acta Metall.* **23**, 553.  
 ESHELBY, J. D. (1954) *J. Appl. Phys.* **25**, 255; (1956) *Solid State Phys.* **3**, 79.  
 FLINN, P. A., AVERBACH, B. L., and COHEN, M. (1953) *Acta Metall.* **1**, 664.  
 FOWLER, R. H. and GUGGENHEIM, E. A. (1940) *Proc. R. Soc. A*, **174**, 189.  
 FREEDMAN, J. F. and NOWICK, A. S. (1958) *Acta Metall.* **6**, 176.  
 FREUNDLICH, H. (1926) *Colloid and Capillary Chemistry* (transl. of 3rd German edition).  
 FRIEDEL, J. (1955) *Phil. Mag.*, **46**, 514.  
 GIBBS, J. W. (1878) *Trans. Conn. Academy* **3**, 439; *Collected Works*, I, p. 219, Yale University Press, New Haven, Connecticut.  
 GLOSSOP, A. B. and PASHLEY, D. W. (1959) *Proc. Roy. Soc. A*, **250**, 132.  
 GUGGENHEIM, E. A. (1948) *Trans. Faraday Soc.* **44**, 1007; (1952) *Mixtures*, Clarendon Press, Oxford.  
 GUINIER, A. (1959) *Solid State Phys.* **9**, 293.  
 GUTTMAN, L. (1956) *Solid State Phys.* **3**, 146.  
 HARDY, A. H. (1953) *Acta Metall.* **1**, 202.  
 HILLERT, M. (1970) *Phase Transformations*, p. 181, American Society for Metals, Metals Park, Ohio.  
 HUME-ROTHERY, W. and POWELL, H. M. (1935) *Z. Kristallogr.* **91**, 23.  
 HUME-ROTHERY, W. and RAYNOR, G. V. (1954) *The Structure of Metals and Alloys*, Institute of Metals, London.  
 JOHANNSON, C. H. and LINDE, J. O. (1936) *Ann. Phys.* **25**, 1.  
 KAUFMAN, L. (1967) *Phase Stability of Metals and Alloys*, p. 125, McGraw-Hill, New York; (1969) *Prog. Mater. Sci.* **14**, 57.  
 KAUFMAN, L. and BERNSTEIN, H. (1970) *Computer Calculation of Phase Diagrams*, Academic Press, New York.  
 KEATING, D. T. and WARREN, B. E. (1951) *J. Appl. Phys.* **22**, 286.  
 KHACHATURYAN, A. G. (1962) *Phys. Met. and Metallg.* **13**, 493; (1973) *Phys. Stat. Sol.* **60**, 9.  
 KIRKWOOD, J. G. (1938) *J. Chem. Phys.* **6**, 70.  
 KURATA, M., KIKUCHI, R., and WATARI, T. (1953) *J. Chem. Phys.* **21**, 434.  
 LANDAU, L. D. and LIFSHITZ, E. M. (1958) *Statistical Physics*, Addison-Wesley, Reading, Mass.  
 LAWSON, A. W. (1947) *J. Chem. Phys.* **15**, 831.  
 LI, Y. Y. (1949) *Phys. Rev.* **76**, 972.  
 LONGUET-HIGGINS, H. C. (1951) *Proc. R. Soc. A*, **205**, 247.  
 MOTT, N. F. and NABARRO, F. R. N. (1940) *Proc. Phys. Soc.* **52**, 86.  
 MULDAWER, L. (1951) *J. Appl. Phys.* **22**, 663.

- MUNSTER, A. and SAGEL, K. (1959) *The Physical Chemistry of Metallic Solutions and Intermetallic Compounds*, paper 2D, HMSO, London.
- ONISAGER, L. (1944) *Phys. Rev.* **65**, 117.
- ORIANI, R. A. (1956) *Acta Metall.* **4**, 15; (1959) *The Physical Chemistry of Metallic Solutions and Intermetallic Compounds*, paper 2A, HMSO, London.
- ORIANI, R. A. and MURPHY, W. K. (1958) *J. Phys. Chem. Solids* **6**, 277.
- PASHLEY, D. W. and PRESLAND, A. E. B. (1958-9) *J. Inst. Metals* **87**, 419.
- PRIGOGINE, I., BELLEMANS, A., and MATHOT, V. (1957) *The Molecular Theory of Solutions*, North-Holland, Amsterdam.
- RHINES, F. N. and NEWKIRK, J. B. (1953) *Trans. Am. Soc. Metals* **45**, 1029.
- RICHARDS, M. J. and CAHN, J. W. (1971) *Acta Metall.* **19**, 1263.
- ROBERTS, B. W. (1954) *Acta Metall.* **2**, 597.
- RUSHBROOKE, G. S. (1938) *Proc. R. Soc. A*, **166**, 296.
- SCHUBERT, K., KIEFER, B., and WILKENS, M. (1954) *Z. Naturforsch.* **9a**, 987.
- SCHUBERT, K., KIEFER, B., WILKENS, M., and HAUFLER, R. (1955) *Z. Metallk.* **46**, 692.
- SHIMOJI, M. and NIWA, K. (1957) *Acta Metall.* **5**, 496.
- SIVERTSEN, J. M. and NICHOLSON, M. E. (1961) *Prog. Mater. Sci.* **9**, 303.
- SMOLUCHOWSKI, R. (1951) *Phase Transformation in Solids*, p. 149, Wiley, New York.
- SWAN, E. and URQUHART, A. R. (1927) *J. Phys. Chem.* **21**, 251.
- SYKES, C. and WILKINSON, H. (1937) *J. Inst. Metals* **61**, 223.
- TAKAGI, Y. (1941) *Proc. Phys. Math. Soc. Japan* **23**, 44.
- THOMSON, J. J. (1888) *Application of Dynamics to Physics and Chemistry*.
- TILLER, W. A., POUND, G. M., and HIRTH, J. P. (1970) *Acta Metall.* **18**, 225.
- TRIVEDI, R. (1970) *Metall. Trans.* **1**, 921.
- WALKER, C. B. (1952) *J. Appl. Phys.* **23**, 118.
- WANNIER, G. H. (1945) *Rev. Mod. Phys.* **17**, 50.
- YANG, C. N. (1945) *J. Chem. Phys.* **13**, 66.
- YANG, C. N., and LI, Y. Y. (1947) *Chin. J. Phys.* **7**, 59.
- ZENER, C. (1951) *J. Appl. Phys.* **22**, 372.

## CHAPTER 7

# *The Theory of Dislocations*

### 28. INTRODUCTION: EDGE AND SCREW DISLOCATIONS

The line defect which we call a dislocation is of great importance in the description of almost all solid-state phenomena. For many years the existence of dislocations was inferred rather than observed, and the theory of dislocations was developed during this time into a reasonably well-ordered body of knowledge. More recently, experimental techniques have been devised by means of which individual dislocations may be seen and their properties studied directly, and the most important advances are now being made in this way.

We cannot hope to cover the theory of dislocations adequately in a single chapter, but we shall summarize the main results. The postulate that real crystals contain dislocations was originally made in order to develop a satisfactory theory of plastic deformation, and the majority of the applications of the theory have been made in this field. We shall not discuss these applications in any detail, but we find it convenient to follow the historical order and introduce the dislocation first by considering the deformation process. This is not strictly necessary, since dislocations can be discussed *ab initio* as line defects possessing certain topological properties (see p. 244), but their glide motion is so important that the deformation approach seems most natural.

Experimental results show that deformation usually occurs by the slipping or gliding of close-packed atomic planes over one another. Before this gliding can begin, the component of the applied shear stress acting across the glide plane and resolved along a close-packed direction of this plane must exceed a certain value. This critical resolved shear stress is characteristic of the state of the material, and is influenced by its thermal and mechanical history. It follows that the atoms in the glide plane maintain a highly ordered (crystallographic) arrangement, and slip does not involve the formation of a locally melted layer, as once supposed. The coupling of the atoms is elastic, not rigid, and it is therefore inconceivable that all the atoms in the glide plane should move simultaneously over the plane beneath. We see then that at a given time, the portion of the crystal on one side of a glide plane may have slipped over the remainder by different amounts in different regions. The simplest definition of a dislocation is that it is a line discontinuity separating two such regions.<sup>†</sup> From the definition it follows at once that the dislocation must either begin and

<sup>†</sup> We give a formal and more complete definition on p. 244.

end on the surface of the crystal or must form a closed line or part of a network in the interior.

A dislocation is characterized principally by its Burgers vector, the scalar magnitude of which is also called the strength of the dislocation. This is the difference in slip, i.e. in relative atomic positions, produced by crossing the dislocation line from one region to another. The Burgers vector  $\mathbf{b}$  is usually written in the form  $[b_1 b_2 b_3]$ , and for cubic structures this may be expressed as  $c[u_1 u_2 u_3]$ , where the  $u_i$  give the crystallographic direction of the displacement, and  $cu_i = b_i$ . Usually, we expect  $\mathbf{b}$  to be equal to the interatomic vector in the glide plane, or at least to a small lattice vector. Dislocations of this kind are called perfect or lattice dislocations. In certain cases, it is also possible to have  $\mathbf{b}$  equal to a fraction of a lattice repeat vector. The discontinuity is then an imperfect or partial dislocation, and the original lattice structure is not preserved when the dislocation line is crossed.

The Burgers vector is constant along any dislocation line, but this invariant characteristic is not sufficient to specify completely the properties of the discontinuity. The strain field, and hence the detailed atomic arrangement, depends on the relation between the Burgers vector and the direction of the line itself. This latter direction need not be fixed; that is, the line may be curved so that its local direction changes continuously in the slip plane. In the early development of the theory, it was assumed that a dislocation line is straight, and that there is a special relation between its own direction and that of its Burgers vector. The assumption of such a relation enables the structure of two fundamental types, edge and screw dislocations, to be discussed in detail. The procedure is useful, since any element of dislocation lying in an arbitrary direction may be resolved into edge and screw components.

In the edge dislocation, introduced by Taylor (1934), the Burgers vector is perpendicular to the line of the dislocation. Such a discontinuity is illustrated in Fig. 7.1. The upper part of the crystal block is gliding over the lower part along the plane  $ABCD$  in the direction  $Ox_1$ . The gliding motion has been completed over the region  $ABPQ$  of the glide plane, and this is separated from the unslipped portion by the dislocation line  $PQ$ . As a result of

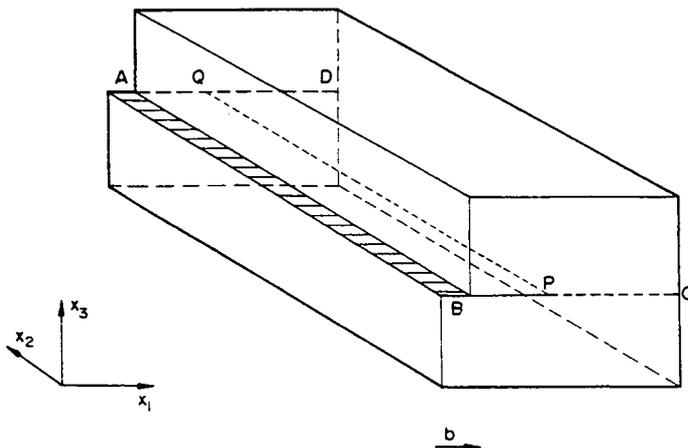


FIG. 7.1. Simple illustration of edge dislocation.

the slip, the atoms are displaced through a distance  $|\mathbf{b}|$  in the direction  $Ox_1$ . It is obvious that in the region of the dislocation line, the atoms on one side of the glide plane are compressed, and those on the other side are extended. This state of strain is characteristic of an edge dislocation.

In the early work, the lattice was assumed to be simple cubic, since this enables the structure of the dislocation line to be readily visualized. For such a lattice, we take  $\mathbf{b} = a[100]$ , so that the block of material shown in Fig. 7.1 has edges parallel to those of the unit cube. If we consider the structure to be made from rows of atoms parallel to  $PQ$ , we see that  $(n+1)$  such rows in the atom plane above the glide plane will be opposite only  $n$  such rows in the atom plane below the glide plane. There is thus an extra plane of atoms in the region of crystal above  $PQ$ . This is illustrated in Fig. 7.2 which shows the approximate atomic

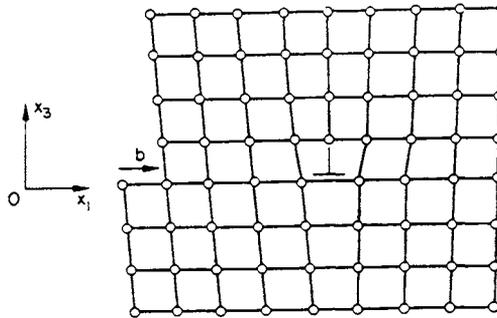


FIG. 7.2. Approximate atomic structure of an edge dislocation in a simple cubic crystal.

structure in a plane perpendicular to  $PQ$ , looking in the direction from  $P$  to  $Q$ . From this figure, we see that instead of the slipping motion discussed above, we could make an edge dislocation by the equivalent hypothetical process of cutting the crystal above  $PQ$  and inserting an extra half-plane of atoms; the dislocation line lies at the edge of this extra half-plane. It is formally equally valid to regard the dislocation as situated at the edge of an extra half-plane of vacant lattice sites inserted below  $PQ$ , i.e. of an extra half-plane of atoms removed from below  $PQ$ . In diagrams, the edge dislocation is conventionally represented by the symbol  $\perp$  or  $\tau$  where the vertical line indicates the extra half-plane of atoms, and the horizontal line the glide plane.

When the crystal structure is not simple cubic, there is an extra half-plane of material of thickness  $|\mathbf{b}|$  on one side of the glide plane, but this does not necessarily consist of a planar arrangement of atoms. Fig. 7.2 no longer shows the atomic arrangement, but it is illustrative of the strain pattern of the edge dislocation; it would be obtained, for example, by scribing a square reference net of lines on the surface prior to introducing the dislocation. It is also clear from this figure that although we consider dislocations to be line defects, each has a finite width and consists of a roughly cylindrical region of bad crystal. A convenient definition of the width of a dislocation is the distance in the slip plane over which the relative displacements of the atoms above and below this plane are more than half their maximum values. The width is determined by the nature of the atomic binding forces, and many

properties of dislocations are sensitive to its exact value. We shall also find in Section 32 that in many close-packed structures each dislocation is dissociated into two parallel (imperfect) dislocation lines, separated by a region of stacking fault. When this happens, the whole dislocation is a planar, rather than a linear defect.

We see from Fig. 7.1 that slip is produced by the migration of the dislocation  $PQ$  from  $BA$  towards  $CD$ . The passage of  $PQ$  right through the crystal from one edge of the slip plane to the other would leave a perfect lattice slipped through a distance equal to the Burgers vector. The formation of a dislocation requires a high energy and will be discussed later; once formed, however, a dislocation in a close-packed structure can move under the action of a small shear stress. This is shown qualitatively by diagrams of the type of Fig. 7.2. If the configurations representing the dislocation in two adjacent atomic positions of the type shown in that figure are compared, it is seen that the movement of the dislocation through this distance has been achieved by displacements of the atoms in the neighbourhood of the dislocation through very small fractions of an interatomic distance. To a first approximation, the energy of the dislocation may be assumed to be constant as this movement is made, so that the dislocation will glide under a vanishingly small shear stress. More accurately we see that when the atomic configuration is symmetrical, as in Fig. 7.2, the dislocation will have a lower energy than it has in an arbitrary position. There is thus a constraining force tending to anchor the dislocation in the lattice; its magnitude depends on the crystal structure, but it is quite small in close-packed metallic structures. The configuration of Fig. 7.2 represents a stable position of the dislocation line; there is also a symmetrical stable configuration halfway between it and the next configuration of the same type. When moving from one stable configuration to another, the dislocation must increase its energy, either by thermal agitation, or because a finite shear stress is applied. The problem is considered further on p. 276.

An edge dislocation can glide only in its slip plane, defined by the dislocation itself and its Burgers vector. It is also geometrically possible for the dislocation to move out of this plane by "climbing" along the perpendicular direction  $Ox_3$  (Fig. 7.1). If the dislocation moves in the  $+x_3$  direction, the extra half-plane of atoms shrinks, or equivalently the extra half-plane of vacancies grows. Such a motion thus requires the removal of atoms from their normal sites into interstitial sites, and their subsequent diffusion away through the lattice, or else the absorption of vacant lattice sites, which diffuse to the dislocation from the surrounding lattice. The reverse is true for motion along  $-x_3$ . The motion of an edge dislocation in a direction out of the slip plane is thus a slow process, depending on the diffusion of vacancies or interstitial atoms.

The absorption or emission of individual point defects implies that the dislocation climbs one atom at a time. Thus the dislocation will not lie entirely in one slip plane but will contain steps where it moves from one such plane to an adjacent plane. These steps are called jogs (Fig. 7.3). We should expect point defects to be absorbed or emitted at the jogs in the dislocation line, the jog moving along the line as this happens. If the local point defect concentration is in excess of the equilibrium value, there will be a decrease in free energy when a defect disappears at the dislocation, this being the binding energy of the defect to the dislocation. Clearly, the absorption of a point defect in a straight region of dislocation

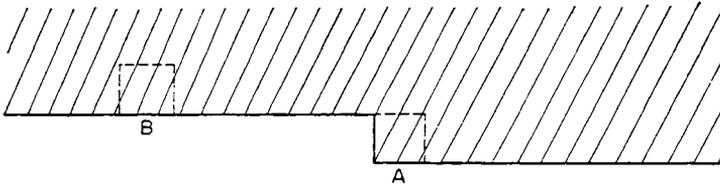


FIG. 7.3. A jog in an edge dislocation. Absorption of a vacancy at the jog *A* does not change the configuration. A higher energy state results from absorption of a vacancy at *B*.

is much less probable, since this would create a new double jog (Fig. 7.3). The energy of this double jog is probably greater than the binding energy, so that if such a configuration forms, it will tend to disappear again. The point defect must thus migrate along the dislocation line until it finds a jog at which it can be absorbed; such movement may be very rapid because of the distorted structure of the line.

The efficiency of a dislocation line as a source or sink for point defects thus depends, at least partially, on the number of jogs which it contains. There will always be a certain jog density maintained in thermal equilibrium, since a dislocation line with a few jogs has a higher entropy to compensate for its higher internal energy. In addition, non-equilibrium jog concentrations may be produced by motion of the dislocation line (see p. 252). We should like to emphasize that this type of dislocation climb is presented here solely as a geometrical possibility; we discuss later the conditions under which it may or may not take place.

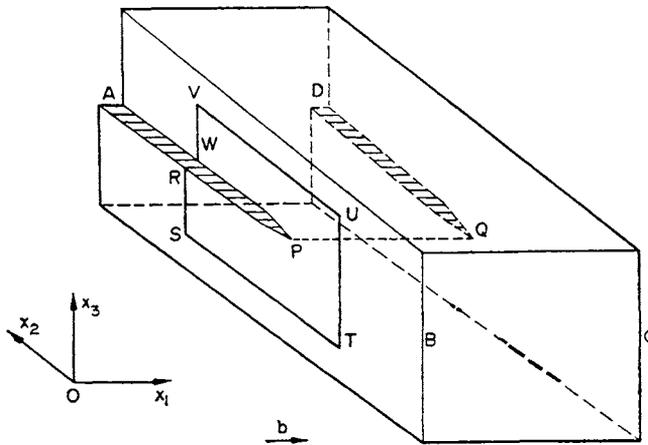


FIG. 7.4. Simple illustration of a screw dislocation.

We now consider the other fundamental type of dislocation, which was introduced by Burgers (1939), and is illustrated in Fig. 7.4. Slip is again occurring along the plane *ABCD* in the direction  $Ox_1$ , and *PQ* is the dislocation line. *PQ* is now parallel to  $Ox_3$ , and slip has been completed over the region *APQD*. Migration of *PQ* from *AD* to *BC* produces the same resultant slip as migration of the dislocation in Fig. 7.1.

The structure of the screw dislocation in the simple cubic structure, or the strain pattern in any structure, is shown by projecting the atomic positions on to the slip plane, as in Fig. 7.5. The filled circles represent atoms immediately below the glide plane, the unfilled circles atoms above the glide plane. The crystal does not contain an extra half-plane of atoms; instead the whole crystal is one continuously connected atomic plane. Consider, for example, the circuit  $RSTUVW$  in Fig. 7.4, which is made entirely in the  $Ox_2x_3$  plane of the crystal. Because of the dislocation, this circuit is not closed; its end-points are displaced by  $\mathbf{b}$ .

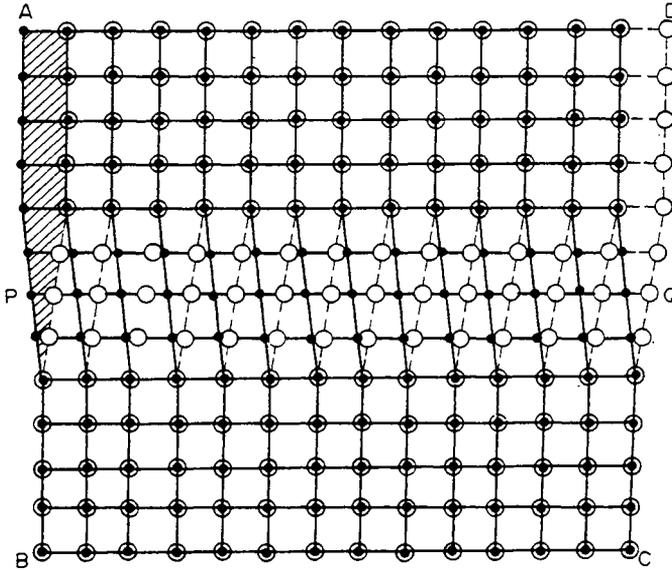


FIG. 7.5. Approximate atomic structure of a screw dislocation in a simple cubic crystal.

A point which continually encircles the dislocation line, following the atomic bonds (i.e. making nearest neighbour jumps) describes a helical path, like the motion of a point on a screw thread.

The dislocation line  $PQ$  of Fig. 7.4 will migrate in the direction  $-Ox_2$  under the action of a small shear stress applied in the direction  $Ox_1$ . For the motion of the screw dislocation, however, there is clearly no distinction between the directions  $Ox_2$  and  $Ox_3$ , and the dislocation can also readily glide in the  $Ox_3$  direction. This is a rapid motion and does not require thermal activation energy. The screw dislocation cannot move slowly by absorption or emission of vacant lattice sites. For the edge dislocation, we saw that the glide plane was the plane containing both the Burgers vector and the dislocation line. If the direction of the dislocation line is given by the unit vector  $\mathbf{i}$ , the glide plane has unit normal  $\mathbf{b} \wedge \mathbf{i} / |\mathbf{b} \wedge \mathbf{i}|$ , and all atomic planes containing  $\mathbf{b}$  are possible glide planes for the screw dislocation.

Figure 7.4 shows that where the screw dislocation emerges at a crystal face there is a step running from the dislocation end to the edge of the face. Movement of the dislocation results in a gradual increase or decrease in the length of this step. In contrast, a step extend-

ing from one edge of a crystal face to another suddenly appears or disappears when an edge dislocation emerges at that face.

The edge dislocation of Fig. 7.1 was obtained by inserting an extra plane of atoms above  $PQ$ . Equally we could have obtained an edge dislocation by inserting an extra plane of atoms below  $PQ$ , and the passage of such a dislocation from  $CD$  to  $BA$  would produce the same resultant slip as the passage of  $PQ$  from  $BA$  to  $CD$ . One description can in fact be converted into the other simply by inverting the crystal. The distinction is less trivial than this implies, since dislocations exert forces on one another. These forces are attractive for dislocations with opposing Burgers vectors and repulsive if the Burgers vectors are parallel. It is convenient to introduce the terms positive and negative dislocations, a positive edge dislocation conventionally having the extra half-plane of atoms in the upper half of the crystal.

Two opposite edge dislocations in the same slip plane are shown in Fig. 7.6. Separation of these lines to the edges of the crystal results in the upper half of the crystal slipping to the right over the lower half. This process is equivalent to moving a single dislocation

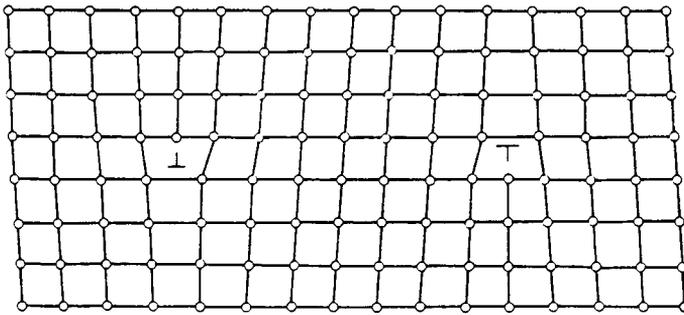


FIG. 7.6. Two opposite edge dislocations on a common slip plane.

through the whole plane. Obviously, if the dislocations move together instead of separating, they will annihilate each other, leaving an unslipped perfect crystal. This results in a lowering of the strain energy, so that the two dislocations attract each other. In the absence of an opposing shear stress, two such neighbouring dislocations will always run together and disappear unless the attractive force is unable to overcome the constraining forces of the lattice. Now consider two opposite dislocations on parallel, adjacent slip planes. There will again be an attractive force, and if the two dislocations run together, they will disappear with the formation of a line of point defects (vacancies or interstitials, depending on whether the extra half-planes do not meet or overlap).

In a similar way, we may form right- and left-hand screw dislocations which are the opposites of each other. The structure of a right-handed screw dislocation is the mirror image of that of a left-handed one. Two such dislocations on a common glide plane can run together, leaving a region of perfect crystal.

The screw and edge dislocations described above are straight lines. One way of forming a more general dislocation in the glide plane is to combine together edge and screw elements. As stated above, a dislocation cannot end within the crystal. The helical surface

of a screw dislocation, for example, can only be terminated in the extra half-plane of an edge dislocation, and vice versa. Thus each time we end a dislocation within the crystal we introduce another dislocation perpendicular to the first but having the same Burgers vector. This is illustrated in Fig. 7.7, which shows a slip plane in which the slipped region

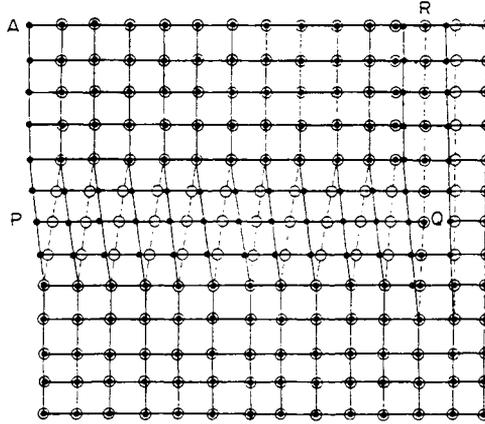


FIG. 7.7. The dislocation line  $PQR$  consists of an edge segment  $QR$  and a screw segment  $PQ$ .

is divided from the unslipped region by screw dislocation segments such as  $PQ$  and edge dislocation segments such as  $QR$ . Note how the extra plane of atoms above  $QR$  terminates in the helical surface wrapped round  $PQ$ .

In Fig. 7.7 we may make the separate elements smaller and smaller, and as the length of each is reduced, the dislocation line approximates more nearly to a curve. If we continue the reduction down to an atomic scale, the separate elements begin to lose their identity, since the width of each becomes comparable to its length. Instead of regarding a small length of curved dislocation line as composed of separate, non-coincident elements with identical Burgers vector, we may regard it as two coincident elements with Burgers vectors parallel and perpendicular to the line (Fig. 7.8). In this mode of resolution, the component edge and screw elements have different Burgers vectors which need not have rational crystallographic directions so long as their sum is a lattice vector. To a certain extent, the manner of resolution is a matter of convenience, but clearly Fig. 7.8 only gives a correct

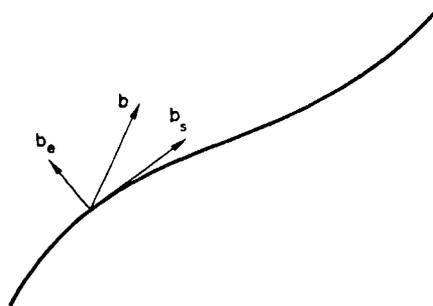


FIG. 7.8. Resolution of the Burgers vector into edge and screw components.

picture if the dislocation line can be represented as a curve, rather than as a zigzag, even on an atomic scale, and if this is possible, the type of resolution shown in Fig. 7.7 is misleading, since the elements are arbitrary. Calculations by Mott and Nabarro (1948) show that a dislocation prefers to be as straight as possible, and the energy rises if the zigzag configuration is adopted. By analogy with surface tension we introduce the idea of the line tension of a dislocation, which may thus be compared to a slightly stretched piece of elastic. The concept of curved, flexible dislocation lines is needed in many applications of the theory.

We see from the above discussion that the general dislocation line in the slip plane is a curve of arbitrary shape, and the edge and screw dislocations described above are best regarded as special orientations of the general line. Each small section of the line may be resolved into edge and screw components; when one of these components is zero, the line is locally a screw or edge dislocation. Of particular interest is the formation of a closed loop of dislocation, isolating a region of local slip. In the absence of a stress, the loop will tend to disappear and leave a perfect crystal. If an external shear stress is applied, a loop of sufficient size will expand, spreading slip over the plane in which it lies.

Although we have only discussed dislocations in a single slip-plane, we have already noted that a dislocation line can move from one slip-plane to another. The most general form of dislocation may have any orientation in the crystal, and Fig. 7.9 shows an elemen-

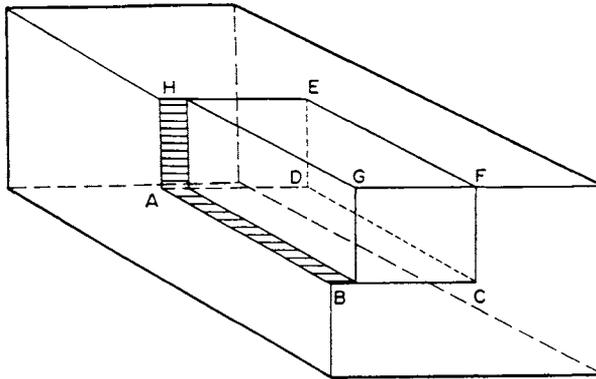


FIG. 7.9. The dislocation line  $CDE$  consists of two perpendicular edge segments.

tary example of a discontinuity in two slip-planes. The region of crystal  $ABCDEFGH$  has slipped through one interatomic distance with respect to the rest of the crystal. This slip has taken place on the planes  $ABCD$  and  $ADEH$ , and its limit is marked by the edge dislocation elements  $CD$  and  $DE$ , making up the complete dislocation line  $CDE$ . By combining perpendicular edge elements of this kind with screw elements in the two slip-planes, we may build up a dislocation line which is curved in three dimensions. More generally, a three-dimensional dislocation may lie in any number of slip planes, which are defined whenever a segment of the line has edge character; these slip planes have a common zone axis which is the direction of the Burgers vector.

The two edge elements of Fig. 7.9 may be formed by making an imaginary cut over the area  $FCDE$  and inserting an extra plane of atoms. Obviously a closed dislocation loop in the  $x_2x_3$  plane can be formed by inserting an extra plane of atoms over any arbitrary area of this plane. The loop will have Burgers vector along  $Ox_1$ , i.e. it is composed entirely of edge elements, and forms the limit of a prismatic region of slipped crystal. The term prismatic dislocation is often used for any closed loop of dislocation line with a Burgers vector inclined to the plane of the loop. A prismatic dislocation of opposite sign could be formed by removing part of an atomic plane (inserting an area of vacancies). An excess concentration of vacancies in a region of perfect crystal could conceivably lower the free energy by condensing to form a spherical cavity or a flat disc of atomic thickness. In the latter case, the disc would collapse inwards to leave a prismatic dislocation loop, and this kind of mechanism was postulated by Seitz (1952) as one way in which dislocations might be introduced into crystals. There is now convincing electron-microscope evidence that dislocations are formed in this way in specimens quenched from high temperatures to produce supersaturations of vacancies.

## 29. GEOMETRICAL PROPERTIES OF DISLOCATIONS

On p. 241, we defined the sign of an edge dislocation. A closed dislocation loop contains edge and screw elements of opposite sign, but it is convenient to adopt a convention in which such a loop may be described by a single Burgers vector. We shall first give a more formal definition of dislocation lines, using the terminology of Frank (1951) already introduced in Chapter 5.

The defects described in Chapter 5 were all of the kind which do not disturb the correlation between the real crystal and the ideal reference crystal. Consider a crystal containing internal strains, so that the mean positions of the atoms are displaced from those of the reference crystal. In any region in which the strains are sufficiently small, there will again be an unambiguous correspondence between the actual atomic positions and the ideal positions, but if the strains become large, this unambiguous correspondence will disappear. When large strains are present, the crystal as a whole can no longer be compared directly with the reference crystal, but for most of the atoms local correspondence will remain. The exceptional regions of "bad crystal" are dislocations.

In a region of good crystal, any four atoms which are arranged in approximately tetrahedral fashion may be associated with four similarly arranged atoms of the reference crystal. A neighbouring atom of the real crystal can then be correspondingly associated with a reference atom, in the sense that its position relative to the other four atoms is almost identical with that of the reference atom with respect to the other four reference atoms. By repeating this process we may trace a path through the real crystal by successive atom jumps and associate with this path a corresponding reference path. The path in the real lattice must be made entirely through good crystal; when such a path closes on itself, it is called a Burgers circuit. If a Burgers circuit encloses regions of bad crystal, the reference path is not necessarily closed, and the displacement needed to close it must be a lattice vector of the reference crystal. This displacement is defined as the resultant Burgers vector of all the dislocation

lines enclosed by the path in the real crystal. An example of a Burgers circuit was shown in Fig. 7.4; if the circuit  $RSTUVWR$  were plotted in the reference crystal, the point corresponding to  $W$  would coincide with the starting point corresponding to  $R$ , and the displacement corresponding to  $RW$  would thus give the Burgers vector. The vector is independent of the starting point of the circuit, and is the same for all circuits which are separated only by good crystal.

Notice that in this definition, the Burgers vector is strictly defined in an ideal reference lattice. Clearly it is equally valid to complete a circuit in the reference crystal, and to define the Burgers vector as the closure failure of the corresponding circuit in the real crystal. The Burgers vector will then connect two lattice points of the real crystal, and since both points are in regions of good crystal, it will differ only very slightly from a lattice vector of the ideal crystal. For most purposes, it is quite immaterial which circuit is regarded as closed, but in the formal theory of continuous distributions of dislocations (p. 316) and similar applications, this point can be important.

The dislocations defined by a closed circuit made entirely through good crystal must be "perfect" since the translations preserve the original structure. Perfect dislocations have Burgers vectors determined entirely by the lattice vectors  $\mathbf{u}$  of eqn. (5.8); the vectors  $\xi$  are not directly related to dislocations, which are a property of the lattice rather than of the crystal structure. Imperfect dislocations are discussed in Section 32. A Burgers circuit which encloses more than one dislocation line may be subdivided by joining points on the circuit through regions of good crystal. Each circuit which can no longer be subdivided encloses a single dislocation.

We may now use this rather formal definition to formulate a sign convention. We choose the positive direction of the dislocation arbitrarily, and describe the Burgers circuit in a clockwise (right-handed) direction when looking along this line. The displacement required to close the reference path then gives both the magnitude and sign of the Burgers vector. Bilby (1951) has given an elementary discussion of this convention, which is sometimes described as  $FS/RH$  (Bilby *et al.*, 1955). This identifies the vector with the displacement from the finish  $F$  to the start  $S$  of the reference path when the right-handed circuit ( $S^2F^1$ ) of the real lattice is closed. The sign of the Burgers vector is changed by taking the closure failure  $SF$ , by making a left-handed circuit, or by closing the reference circuit and measuring  $F^1S^1$ . Thus  $FS/RH = SF/LH = F^1S^1/LH = S^1F^1/RH = -F^1S^1/RH$ , etc. There is no general agreement on which convention should be used.

If we have a closed dislocation loop, we may define the positive direction of the line by specifying that the loop is to be traversed in a certain sense, and we then have the same Burgers vector at all points on the loop. Consider a loop lying in a single glide plane. At two points along the loop, the strain field will be of the type we previously associated with positive and negative edge dislocations, and at two other points the local arrangement will correspond to right- and left-handed screw dislocations. Whereas we should previously have stated that the two edge components have equal and opposite Burgers vectors, we should now state that they have the same Burgers vector but opposite directions. The two screw elements are still describable as right- and left-handed (simple rotation can never convert a right-handed thread into a left-handed thread) but are now regarded as having opposite

directions, so that they possess the same Burgers vector. We thus see that there is always an ambiguity in the sign of the Burgers vector until the direction of the line itself is specified in some way. For three-dimensional dislocations, or closed two-dimensional loops, it is often more convenient to regard elements of the line which are opposites in the sense that they can mutually annihilate as having opposite directions rather than opposite Burgers vectors.

Suppose the Burgers circuit in the real crystal encloses two separate regions of bad crystal. If we displace the circuit in the general direction of these dislocations (i.e. along the dislocations), it is geometrically possible for the two bad regions to coalesce into a single bad region. Two perfect dislocations may thus unite to form a single dislocation or, conversely, a dislocation may split into two dislocations in the interior of the crystal. Since the Burgers circuits enclosing the dislocation or its branched pair are entirely in good crystal, the resultant Burgers vector is unchanged and

$$\mathbf{b}_1 + \mathbf{b}_2 = \mathbf{b}_3. \quad (29.1)$$

The meeting of three or more dislocation lines in the interior of a crystal is called a node. Clearly, in a node of three lines we are free to regard any two of them as resulting from the decomposition of the third. The relation between their Burgers vectors is then expressed more symmetrically if we define the positive direction of each line as the direction looking outward from the node. By describing right-handed Burgers circuits round each of the lines, we then find

$$\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3 = 0,$$

or more generally if  $i$  dislocation lines meet at a node,

$$\sum_i \mathbf{b}_i = 0. \quad (29.2)$$

Equation (29.1) may also be considered as a dislocation reaction in which (by analogy with chemical reactions) the two dislocations  $\mathbf{b}_1$  and  $\mathbf{b}_2$  combine to form a new dislocation  $\mathbf{b}_3$ . In theory, any number of dislocations may contribute to a reaction. Non-parallel dislocations in the same glide plane may glide together, uniting over part of their lengths, at the ends of which nodes are formed. The positions of the nodes then change continuously as the reaction proceeds.

Consider any element of dislocation line, lying along a direction specified by the unit vector  $\mathbf{i}$  and having Burgers vector  $\mathbf{b}$ . If we resolve into edge and screw components, as described on p. 242, the edge component of the Burgers vector lies in the slip plane specified by  $\mathbf{b}$  and  $\mathbf{i}$ , and is perpendicular to  $\mathbf{i}$ , and the screw component is parallel to  $\mathbf{i}$ . The screw component is thus given by

$$\mathbf{b}_s = (\mathbf{b} \cdot \mathbf{i})\mathbf{i} \quad (29.3)$$

and the edge component by

$$\mathbf{b}_e = \mathbf{b} - \mathbf{b}_s = (\mathbf{b} \wedge \mathbf{i}) \wedge \mathbf{i}. \quad (29.4)$$

We introduced a dislocation as a line discontinuity in the glide plane, but we have already seen that for a general dislocation loop there is no unique glide plane. However, if we draw generators from the dislocation parallel to the Burgers vector, we obtain a prismatic

surface which may be called the virtual glide surface for the dislocation (Read and Shockley, 1952). From the previous consideration of slipping movements, it follows that any movements of parts of the dislocation loop along this surface are slipping motions; such motions have the property that the projected area of the loop on any plane normal to  $\mathbf{b}$  remains unchanged. Conversely, any motion of the loop which changes this projected area requires the movement of edge elements out of their slip planes, and is a diffusive or climbing motion. It is also obvious that an element of loop which has pure screw character may move off the virtual glide surface without diffusion, so long as this does not change the projected area.

We should also note which kinds of motion are likely under the action of applied stress. In Fig. 7.10 the virtual glide surface has been drawn as a cylinder for simplicity. The ini-

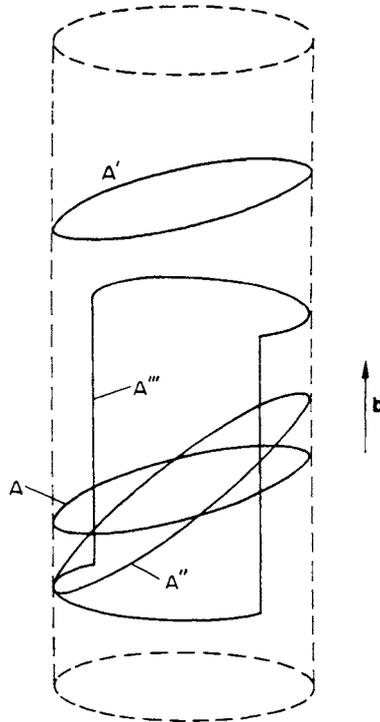


FIG. 7.10. Possible movements of a dislocation loop on its virtual glide surface. An external shear stress may displace the loop  $A$  to positions such as  $A''$  and  $A'''$ .

tial dislocation loop  $A$  lying on this surface may be imagined to move in two ways: bodily along the surface without change in length, or by gliding to a new position  $A''$  in which its length is greatly increased or decreased. Slipping motions of the second kind are found when the loop is subjected to a shear stress, and are the three-dimensional analogues of the simple expansion of a glide plane loop described on p. 245. In slipping motions of the first kind there is a displacement of the material inside the virtual glide surface relative to that outside the surface. Such a motion cannot relieve any externally applied shear stress,

but it may be responsible for the phenomenon observed in prismatic punching experiments. The possible dislocation motions just discussed refer only to perfect dislocations; partial dislocations are subjected to additional constraints and have more restricted possibilities of motion.

Early observations of the surface of deformed materials by electron-microscopy indicated that in some circumstances large amounts of slip may occur on a single glide-plane. These slip steps represent dislocations which have escaped from the material, but it is also well established that the density of dislocation lines inside the crystal (see p. 313) increases by several powers of ten during work hardening. We thus require a mechanism for the production of large numbers of perfect dislocations during plastic deformation, and we shall prove later that such dislocations cannot be nucleated spontaneously by the combined effects of the external stress and the thermal energy of the lattice. An important development in the theory of dislocations was the recognition by Frank and Read (1950) that a process which is in principle purely geometrical can be used to produce new dislocations from old dislocations.

Consider a crystal which contains a dislocation line  $CDE$  composed of two edge elements, exactly as in Fig. 7.9. Suppose that when an external shear stress is applied to the crystal, the part  $DE$  is prevented from gliding in some way, so that the horizontal planes of the figure are the only active glide planes. Under the action of a suitable stress, the portion of the line  $CD$  will move in its glide plane, and as it does so the crystal above this plane glides over the crystal below it. The point  $D$  must remain fixed, so the line  $CD$  rotates about  $D$ . The crystal is slipped through one Burgers vector for each complete rotation. Actually,  $CD$  will not remain radial, since the parts of the line  $CD$  will glide initially with constant linear velocity (except very close to  $D$ ). The angular velocity of the interior parts of the line will thus be greater than that of the outer parts, and the line will wind up into a spiral shape. Once an equilibrium spiral has been formed, it will continue to rotate with constant angular velocity, increasing the slip in each region of the glide plane by  $\mathbf{b}$  for each complete revolution. We shall not calculate the shape of this equilibrium spiral, but an analogous problem in the theory of crystal growth will be treated in Part II, Chapter 13.

In the example just discussed, the dislocation line  $CDE$  only leaves the active glide plane at one point  $D$ , and  $CD$  extends to the surface of the crystal. This is improbable, even for the low dislocation densities of a well-annealed crystal, unless the whole configuration is close to the surface. As we shall describe later, the dislocations in an annealed crystal are largely concentrated into two-dimensional networks (sub-boundaries), with a few dislocations forming a three-dimensional network inside the subgrains. In the interior of a crystal we thus expect a limited length of any one dislocation line to lie in a given glide plane, the line leaving the glide plane at two points. This gives the geometrical configuration necessary for the operation of a double-ended Frank-Read source as opposed to the single-ended source described above.

We suppose that the dislocation line is anchored at the points of emergence from the slip plane, but is free to glide in this plane. Under the action of an applied stress, it expands in the glide plane, forming a loop which eventually winds round and joins together again. When this happens, a closed loop of dislocation is formed, and this spreads out over the

glide-plane, whilst the original length of line repeats the process. This action is illustrated at various stages of the expansion in Fig. 7.11. The extra energy of the dislocation line which has been created has to come from the work done by the external stress. The operation of the source is opposed by the line tension of the dislocation, which tends to straighten it, and the critical stage is that shown at (c) in Fig. 7.11 where the dislocation line reaches its minimum radius of curvature. The greater the distance between the fixed points *A* and *B*, the lower the stress needed to activate the source. A single-ended source should become active at about half the stress needed for a double-ended source of the same length, since only one end of the gliding dislocation segment is anchored.

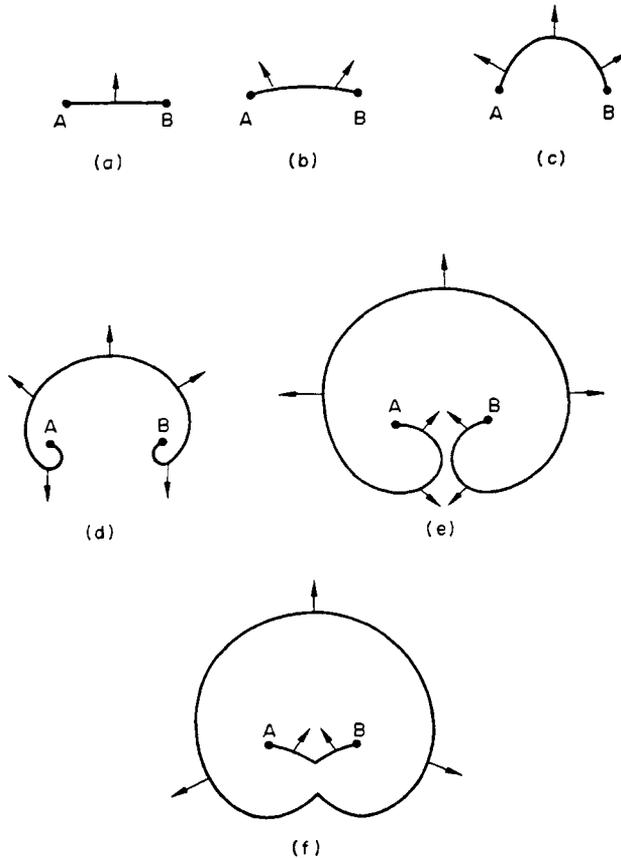


FIG. 7.11. Successive stages in the operation of a double-ended Frank-Read source.

The points *A* and *B* of Fig. 7.11 could represent places where the dislocation line merely bends out of the glide plane, as in Fig. 7.9, but it is perhaps more likely that nodes are formed at these places. The geometrical action of the source is unchanged provided the new dislocations formed at the nodes do not lie in the glide plane but have Burgers vectors which do. Nodes may be a method of ensuring firm anchoring of the ends of the source,

since the dislocations produced by the dissociation of the glide dislocations may be types which cannot glide readily.

The theory of Frank–Read sources received early experimental support from the work of Dash (1956), who revealed such a configuration in a silicon crystal by precipitating copper on the dislocation lines which were then visible when the crystal was examined in infrared light. Observations of source operation are rather rare, however, and are mainly confined to very lightly deformed crystals (e.g. Miltat and Bowen, 1970). Whilst Frank–Read sources are undoubtedly operative in certain circumstances, there is also abundant evidence that in some materials glide dislocations originate from inclusions, from boundaries, or from sources near a free surface. Dislocation multiplication also often occurs by spreading of slip from one plane to a small group of neighbouring planes. This can happen by cross-slipping of screw segments or by interaction with point defects to form “super-jogs”; in either case, the part of the dislocation line which is displaced acts as a source, sending out a few dislocations on a parallel plane, before the cross-slip or glide again takes place. This dynamic modification of the Frank–Read mechanism is often described as a double cross-slip source.

We now turn to consider an interesting topological property of a double-ended source in which nodes are formed at the points of emergence. There is no necessity for all the dislocations at a node to have Burgers vectors lying in the glide plane so long as eqn. (29.2) is satisfied, and we now consider what happens if they do not. If the Burgers vectors of (say) two emerging dislocations have (equal and opposite) components normal to the glide plane, the rotating dislocation is displaced upwards or downwards through a distance equal to this component for each complete revolution. This could lead to uniform slip on a series of parallel glide planes if the rotating dislocation is perfect, but for our purpose, the more interesting application is to mechanical twinning and martensitic transformations, when this dislocation is imperfect. We shall give a detailed description of this modified form of Frank–Read source, originally suggested by Cottrell and Bilby (1951) in Section 32 and Part II, Chapter 20.

Another modification of the Frank–Read source mechanism is its application to dislocation climb. In Fig. 7.11 we now assume that the dislocation  $AB$  is a pure edge dislocation with its Burgers vector normal to the plane of the paper. If there is a large excess of vacancies in the region of the dislocation, it can climb by addition of these vacancies, and in so doing it increases its length and takes up a curved shape. The successive curves of Fig. 7.11 are now successive stages in the climb of the dislocation line, and eventually a closed loop of prismatic dislocation is formed, together with an edge element which can repeat the process. The closed loop at the edge of a disc of vacancies can grow outwards, removing an atom plane from the crystal as it does so. The configuration can thus act as a continuous sink for vacancies (source of interstitials), or by climbing in the opposite direction as a source of vacancies (sink for interstitials). This is usually described as a Bardeen–Herring source (Bardeen and Herring, 1952), and was first applied to the Kirkendall effect in diffusion, (see p. 403).

Figure 7.12 illustrates the climb of a dislocation line fixed at  $A$  and  $B$  which is of the mixed type. The Burgers vector is parallel to  $A'A$ , so that the glide plane is  $ABA'$  and the configu-

ration could act as an ordinary Frank-Read source in this plane. Suppose that the dislocation climbs. The line  $AB$  then becomes curved and lies along the surface of some cylinder with generators parallel to  $\mathbf{b}$ ; this surface is now the new virtual glide surface for the line  $AB$ . The area  $F$  between  $A'B$  and the new projection of the dislocation line on the plane

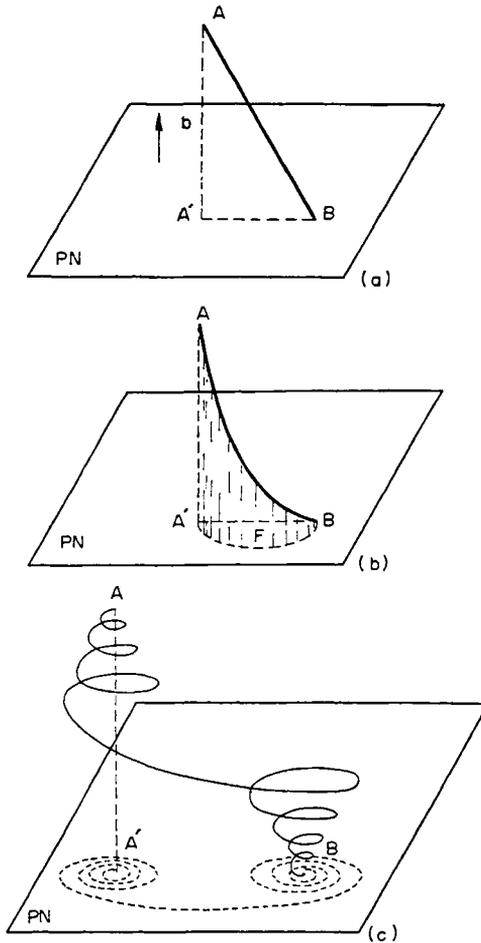


FIG. 7.12. To illustrate the climb of a general dislocation line  $AB$  into a spiral configuration (after Amelinckx *et al.*, 1957).

normal to  $\mathbf{b}$  specifies the amount of material added or removed (i.e. the number of vacancies emitted or absorbed). Further climb displaces the dislocation line in a direction normal to the virtual glide surface at each point, and the whole line may change into a single spiral, or into a double spiral as shown in Fig. 7.12(c). The successive turns of the spiral (which has a radius of curvature decreasing towards the fixed points) will tend to repel each other by glide along the virtual glide surface. If  $A'A$  is large in comparison with  $A'B$  and with the spiral radius, the outer turns of the spiral will glide towards the centre, and a simple

helix will be formed. Two such helices meeting in opposite directions produce a series of closed loops of dislocation line.

In the limit, helical dislocations can be produced by the interaction of point defects and pure screw dislocations. The point defects collapse in discs to give pure edge loops normal to the screw dislocation line, and the interaction of the screw with these prismatic loops then gives a helix. The earlier conclusion that pure screw dislocations cannot climb thus needs modification. Dislocation lines of the general form just discussed have been called spiral prismatic dislocations (Seitz, 1952).

It is necessary to make the same kind of reservation about Bardeen–Herring sources as we made about Frank–Read sources. Experimental evidence on dislocation climb appears rather confusing, insofar as there are well established cases in which dislocations do move by absorption or emission of point defects, and other equally well authenticated circumstances in which dislocations are ineffective as sources or sinks. Kuhlmann-Wilsdorf *et al.* (1962) suggest that all these experiments can be rationalized by making a distinction between old dislocations and new dislocations. New dislocations have been formed or have moved at fairly low temperatures, and have not been subjected to subsequent ageing treatment. Old dislocations, in contrast, have been stationary in the structure whilst it has been heated to moderately high temperature. The evidence suggests that new dislocations can climb, but old dislocations cannot do so, at least in f.c.c. structures. A possible reason for this is that the jogs are poisoned by impurity atoms after heating to temperatures at which such atoms can move.

The final geometrical property to be considered is the effect of two dislocation lines, lying in different planes, which glide through each other. Suppose we have two dislocations,  $A$  and  $B$ , in intersecting planes, and that  $A$  glides and so cuts through  $B$ . The motion moves the part of the crystal on one side of the glide plane of  $A$  by the vector  $\mathbf{b}_A$  relative to the part on the other side. The two parts of the dislocation line  $B$  separated by the glide plane of  $A$  must therefore suffer this relative displacement, and since dislocations cannot end within the crystal, the two original parts of  $B$  are joined by a short length of line (i.e. by a jog), equal in length and magnitude to  $\mathbf{b}_A$ . Equivalent arguments apply to the line  $A$  which acquires a jog equal to  $\mathbf{b}_B$ , since we could equally well suppose that  $B$  moved through  $A$ . Thus whenever two dislocations move through each other, they both acquire a jog.

Some simple geometrical configurations which arise when two lines are of simple types and are perpendicular to each other are of interest. A gliding dislocation which acquires a jog can always continue to move easily if its slip plane contains the line of the jog, i.e. if it contains the Burgers vector of any dislocation which it has intersected. If this is not true, only pure edge dislocations can continue to move conservatively together with a fixed jog produced by intersection. When both dislocations are of edge type, the two jogs produced may both be pure screws, and can be eliminated completely by the continuing glide motion, or else one dislocation may acquire an edge jog which continues to glide readily, whilst the other acquires a jog parallel to itself (i.e. increases its length slightly). If an edge dislocation glides through a screw, it acquires an edge jog which can continue to glide, and the screw also acquires an edge jog. Two screw dislocations gliding through each other both acquire edge jogs.

An edge jog in a pure screw dislocation is able to glide only along the dislocation. Thus if the screw dislocation itself is moved in any direction (other than that parallel to the jog), and the jog is carried along with it, a line of vacant lattice sites or interstitials must be left behind by the climb of the edge jog. Seitz (1952) first suggested that during plastic deformation there is a large increase in the concentration of point defects because of this and similar processes. The generation of such point defects would greatly hinder the movement of a jogged screw dislocation.

A jog which has edge character in a general dislocation line is also unable to move conservatively if it is fixed in position along the dislocation line. Thus motion of the line will also lead to the generation of point defects, though in lower density if the jog is not pure edge. However, Seeger (1955a) pointed out that the jog can move conservatively in a slip plane defined by its own direction and that of the Burgers vector, that is, in a direction inclined to the normal to the dislocation line. This should be possible if the jog moves along the line as the line itself moves forward. Although occasional point defects might still be produced, for example, if the jog is temporarily halted for some reason, this type of conservative motion should be much more probable than the non-conservative motion.

The possibility of conservative jog motion may reduce the importance of moving jogs as sources of point defects, but there are numerous other ways in which defects may be created by moving dislocations. Figure 7.13 illustrates the production of an edge dislocation dipole

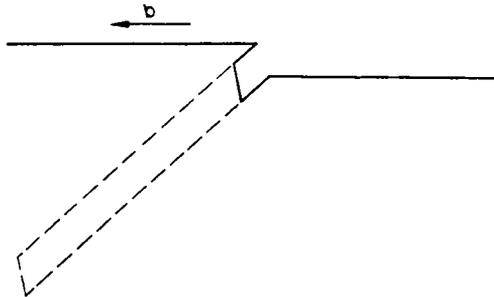


FIG. 7.13. Production of point defects by "pinching off" of edge dislocation dipole.

when a jog in a pure screw is held back by the effect just discussed, the rest of the dislocation continuing to glide. If the jog is only one or two Burgers vectors in length, the pair of edge dislocations are equivalent to a line of point defects, which may be pinched off by the linking together again of the two screw parts of the dislocation line. If the jog is rather longer than this, eventual pinching off will leave a loop of prismatic dislocation. Electron-microscope observations have shown that at sufficiently high temperatures this will break up into a number of approximately circular dislocation loops, and these will eventually disappear by climb. Finally, if the jog is a long "super-jog", produced by cross-slip of part of the original screw, the parts of the dislocation line may behave independently, and may act as single-ended dislocation sources by spiralling around the super-jog. These effects have all been observed in the electron-microscope (e.g. Low and Turkalo, 1962), and there is increasing evidence that the production of "debris" in the form of dislocation dipoles is important in the deformation of many materials.

In crystals where the dislocations are extended (see p. 287), the equilibrium structure of a jog is very difficult to calculate. Detailed consideration of the possible configurations leads to the conclusion (Hirsch, 1962) that vacancy-producing jogs may not be able to move sideways along the dislocation line, and so cannot glide conservatively with a dislocation. Conservative motion of interstitial producing jogs is shown to be much easier, although this motion is thermally activated and may require an increasing stress at very low temperatures. Whatever the mechanism, there is good evidence for some metals that vacancies but not interstitials are produced in considerable numbers during plastic deformation. This is important in some low-temperature transformation processes.

### 30. DISLOCATIONS IN AN ELASTIC MEDIUM

No real metals have simple cubic structures, and some extra factors have to be considered when applying the results of the last section. Before doing this, we shall briefly review some of the more mathematical aspects of the theory. The results of this section are independent of the actual crystal structure, and are, indeed, obtained by ignoring the crystalline nature of the material.

When a dislocation is present in an external stress field, the energy of the whole assembly may be reduced by movement of the dislocation line. Suppose a small element of the line is displaced in the direction  $j$ . We define the force acting on this element in the  $j$  direction as the rate of change of energy for movement in this direction. The force acts on the dislocation line as a configuration, and not on the atoms which constitute the core of the dislocation at any one instant; care must be taken not to extend the concept beyond the formal definition. The difficulties of precisely defining the force on a dislocation have been discussed by Peach (1951).

For simplicity, consider first a simple edge or screw dislocation (Fig. 7.1 or 7.4) with an applied stress  $X_{31}$  across the  $x_1x_2$  plane in the  $x_1$  direction. If the dimensions of the crystal in the  $x_1$  and  $x_2$  directions are  $L_1$  and  $L_2$  respectively, the applied force is  $X_{31}L_1L_2$ . By moving the dislocation across the whole slip plane, the two halves of the crystal are displaced a relative distance  $b$  in the  $x_1$  direction, so that the work done is  $bX_{31}L_1L_2$ . If the force on the dislocation line has magnitude  $f$  per unit length, the total force on the edge dislocation is  $fL_2$ , and in moving it across the crystal the work done is  $fL_2L_1$ . Similarly, the screw dislocation has a force  $fL_1$  acting on it, and the work done is again  $fL_1L_2$ . Equating the expressions for work done, we find for both edge and screw dislocations

$$f = bX_{31}. \quad (30.1)$$

The force is a vector perpendicular to the length of the line, i.e. in the  $Ox_1$ ,  $Ox_2$  directions for edges and screws respectively.

The above derivation is obviously approximate; the general expression was first given by Peach and Koehler (1950). Consider any element of dislocation line  $dd$  present in an external stress field. The field is completely represented by the symmetrical stress tensor  $\mathbf{X}$ , the components of which are functions of position. The forces acting across an arbitrary plane at any point in the stress field are  $\mathbf{X}d\mathbf{O}$ , where  $d\mathbf{O}$  is the vector area of the plane.

Now let the stress field move the dislocation line a distance  $dx$ , where the direction of  $dx$  specifies the direction of the resultant force on  $dd$ . The area swept out by  $dd$  during the movement is  $dx \wedge dd$ , and hence the force acting across this area is  $X(dx \wedge dd)$ . As a result of the motion, the two parts of the crystal on either side of the swept area are displaced relative to each other by the Burgers vector  $b$ . Hence the work done during the displacement is  $b \cdot X(dx \wedge dd)$ , and since  $X$  is a symmetric tensor, this may also be written  $Xb \cdot (dx \wedge dd)$ . If  $f$  is the force per unit length acting on the element of dislocation line during the displacement, it does work  $|dd| f \cdot dx$ . Equating the two expressions, we find

$$f = Xb \cdot i, \tag{30.2}$$

where  $i$  is the unit vector in the direction of the dislocation element ( $dd$ ) at the point considered.

Peach and Koehler emphasised that this result is perfectly general, and especially that the stress field  $X$  may arise in any way from externally applied forces or from interactions with dislocations and other irregularities in isotropic or anisotropic crystals. However, Weertman (1965) pointed out that eqn. (30.2) appears to give wrong results in certain cases, and he suggested that  $X$  should be replaced by the "deviator" stress  $X'$ , where

$$X'_{ij} = X_{ij} - \left(\frac{1}{3}\right)\delta_{ij}X_{kk}.$$

The difficulties arise because the division of forces into mechanical, chemical, and other terms is not necessarily unique when climb processes are involved. According to Lewthwaite (1966) when a dislocation loop is formed in the presence of an applied stress  $X$  in a body in which material is conserved, the change in energy is given by

$$W_d = - \int_O Xb \cdot dS - \int_A Xw \cdot dA, \tag{30.3}$$

where  $O$  is the cut surface used to introduce the dislocation,  $A$  is the external surface, and  $w$  is a displacement of that surface caused by rearrangement of material after the dislocation is created. A virtual displacement of the dislocation element  $dd$  in any direction  $dx$  now gives

$$\delta W_d = - Xb(dx \wedge dd) - \int_A X \delta w \cdot dA. \tag{30.4}$$

If this equation is used to define the force per unit length  $f$  on the element  $dd$  through the identity

$$\delta W_d = |dd| f \cdot dx,$$

it follows on equating the two expressions that the first term in (30.4) gives the Peach-Koehler formula (30.2), whilst the second term represents modifications caused by material transfer, and can only be given explicitly when the boundary conditions for this transfer have been specified. Thus the climb of an edge dislocation in a plane normal to the end faces of a cylindrical crystal subjected to a uniform tensile or compressive stress  $X_{33}$ , involves the emission or absorption of vacancies at the dislocation line, and the effective force depends on the sink or source of vacancies. For vacancies absorbed or emitted uniformly at the

external surface of the crystal, the force has magnitude  $f = 2bX'_{33}/3 = bX'_{33}$ , whilst if the vacancies all appear or disappear at the cylindrical surfaces,  $f = bX_{33}$ . If the vacancies are retained within the lattice, the force is  $bX_{33}(1 - (v - v_{\square})/3v)$  where  $v_{\square}$  is the volume of a vacancy. Weertman wrote this last expression as  $bX'_{33}$  by incorporating the term in  $v_{\square}$  in the chemical potential of the lattice.

The above equations were derived on the assumption that the dislocation moves in the direction of the resultant force. Nabarro has pointed out that we are usually interested only in the effective force causing conservative motion in the glide plane, and has shown how this leads to the law of constant resolved shear stress mentioned on p. 235. The normal to the glide plane is the unit vector  $\mathbf{n}$  defined by

$$\mathbf{n} = (\mathbf{b} \wedge \mathbf{i}) / |\mathbf{b} \wedge \mathbf{i}|$$

The force  $\mathbf{f}$  may be resolved into a component  $\mathbf{f}_g$  in the glide plane and a component  $(\mathbf{f} \cdot \mathbf{n})\mathbf{n}$  normal to the glide plane, where

$$\mathbf{f}_g = \mathbf{f} - (\mathbf{f} \cdot \mathbf{n})\mathbf{n} = -\mathbf{n} \wedge (\mathbf{n} \wedge \mathbf{f})$$

using a standard equation for a vector triple product. Substituting for  $\mathbf{f}$  from eqn. (30.2),

$$\begin{aligned} \mathbf{n} \wedge \mathbf{f} &= \mathbf{n} \wedge (\mathbf{X}\mathbf{b} \wedge \mathbf{i}) = (\mathbf{n} \cdot \mathbf{i})\mathbf{X}\mathbf{b} - (\mathbf{X}\mathbf{b} \cdot \mathbf{n})\mathbf{i} \\ &= -(\mathbf{X}\mathbf{b} \cdot \mathbf{n})\mathbf{i} \end{aligned}$$

and, finally,

$$\mathbf{f}_g = (\mathbf{X}\mathbf{b} \cdot \mathbf{n})\mathbf{n} \wedge \mathbf{i} = (\mathbf{X}\mathbf{b} \cdot \mathbf{n})\mathbf{j}, \quad (30.5)$$

where  $\mathbf{j}$  is the unit vector in the glide plane perpendicular to  $\mathbf{i}$ . The force across unit area of the glide plane is  $\mathbf{X}\mathbf{n}$ , and hence the force per unit length of dislocation tending to produce glide,  $\mathbf{X}\mathbf{b} \cdot \mathbf{n} = \mathbf{X}\mathbf{n} \cdot \mathbf{b}$ , is equal to  $|\mathbf{b}|$  times the component of the applied stress across the glide plane resolved in the direction of  $\mathbf{b}$ . If dislocations begin to move at a critical value of  $\mathbf{f}_g$ , or if their mean velocity is determined by  $\mathbf{f}_g$  alone, eqn. (30.5) provides a theoretical justification for the experimental Schmid law that yielding begins at a critical resolved shear stress.

In some crystalline materials, e.g. b.c.c. metals at low temperatures, the Schmid law is not observed, and it must be concluded that the force in the glide plane is not the only factor controlling motion of the dislocation line. Moreover, there is an asymmetry in the observed behaviour such that reversal of all the components of  $\mathbf{X}$  does not necessarily lead to a reversal of the motion. Not only is the critical stress for slip different in tension and compression, but the operative glide-plane of the crystal may change when the stress is reversed. Such effects, although best documented in b.c.c. metals (Christian, 1970), may well exist in other materials in which the core structure of the dislocation is an important factor in determining its mobility (see Section 31).

A single dislocation is highly mobile, and in the presence of a stress field will tend to move in the direction of  $\mathbf{f}_g$ . In problems of interest, the stress field includes contributions from other dislocations in the structure, each of which is the centre of an internal stress field. In order to calculate the stress distribution, we may disregard the atomic structure of the material, and treat it as an elastic continuum; this procedure is justified so long as we are interested in stresses which vary only slowly over distances of atomic dimensions.

We may form a dislocation in a continuum in the following way. A slit is cut over part of a plane, and the two surfaces of the cut are moved over each other a distance  $\mathbf{b}$  and then rejoined; in order to do this, it may be necessary to add or remove thin layers of material over the cut surfaces. If  $\mathbf{b}$  is perpendicular to the edge of the slit, this edge forms an edge dislocation; if  $\mathbf{b}$  is parallel to the edge of the slit, it forms a screw dislocation. In a true continuum, the stresses would be infinite at the dislocation line, and we have therefore to imagine a narrow cylinder of material a few atomic spacings in diameter cut out along the edge. Physically, the significance of this is that the condition of slowly varying stresses is no longer satisfied near the dislocation line, and the continuum approximation is invalid there. The actual displacements are finite everywhere, but in the region of bad crystal, the stresses vary too rapidly with distance to be treated as continuous.

A general dislocation loop may be formed in a continuum in the same manner. We cut the material over any surface which has as its limit the dislocation line. The two cut surfaces are then displaced by the vector  $\mathbf{b}$  and rejoined. As before, it is necessary to remove a thin cylinder of material along the dislocation core, and also to remove or add thin layers of material wherever the displacement is not parallel to the surface. The dislocation is determined entirely by the limiting line of the cut surface and the vector displacement; the shape of the cut surface is without significance.

The removal of the thin cylinder along the dislocation core signifies mathematically that we can only introduce dislocations into a multiply connected body. The theory of dislocations in an elastic body was formulated long before they were introduced into the physical theory of crystals, and dislocations of more general types were considered. Consider a doubly connected body obtained by boring a cylindrical hole through a solid body. Now cut the body over any surface which terminates on the hole, as in the above procedure, and give the two cut surfaces any elastic deformation and displacements, adding or removing material as necessary, before rejoining them. We have now produced the most general type of dislocation, known as a Somigliana dislocation. A more restricted class of elastic dislocations is obtained if we only permit rigid displacements of the cut surfaces, and these are known as Volterra dislocations. There are six fundamental types of Volterra dislocation, two being (equivalent) edge dislocations and one a screw dislocation obtained by vector translations of the cut surfaces normal or parallel to the edge of the cut. The other three types of Volterra dislocation correspond to rotations of the two cut surfaces with respect to each other, and do not represent single dislocations in crystals.

Following a suggestion by Frank, the rotational Volterra dislocations are now termed "disclinations" (Nabarro, 1967). A dislocation line is characterized by the net displacement of a point when carried around the line, and a disclination line is similarly characterized by the net rotation of a vector carried around a circuit enclosing the line. Clearly, if the lattice structure is to be preserved except in the immediate vicinity of the line, the rotation angle must be  $2\pi/n$  when the axis of rotation is an  $n$ -fold symmetry axis. Lattice disclinations must thus have large strengths, and hence do not occur in three-dimensional crystals since the elastic strains are large at large distances from the singularity. However, disclinations occur in liquid crystals (Kléman and Friedel, 1969), and it is also possible to envisage disclinations of small strength in internally stressed crystals of finite size. In the latter case,

the disclination must form the edge of a planar surface across which there is a rotational discontinuity which may now be made indefinitely small; by analogy with the corresponding (translational) dislocation (see p. 286), this configuration may be called a partial disclination. An example is a low-angle grain boundary, which terminates inside the crystal, so that its edge is topologically a disclination. Such a boundary may be regarded as an array of line dislocations (see Chapter 8), and a disclination is thus a discontinuity in the regular dislocation array.

The three translational Volterra dislocations are characterized by constant Burgers vectors, and hence correspond to real dislocations in crystals. The more general dislocations have varying Burgers vectors, and hence can be used to represent any arbitrary collection of real crystal dislocations. Any network of dislocation lines can then be regarded as a suitable Somigliana dislocation, and this has been useful in some problems. Centres of compression or dilatation (point defects) may also be regarded as Somigliana dislocations. We shall generally use the word dislocation to mean only the translational Volterra dislocations of constant Burgers vector, although we shall also introduce the idea of a "surface dislocation" in Section 34.

To calculate the stress field of an edge dislocation, we suppose that it lies along the axis of a long cylinder of infinite radius. We take the  $\mathbf{b}$  direction as  $Ox_1$  and the axis of the cylinder as  $Ox_3$ . Over most of the cylinder, the displacements in the  $Ox_3$  direction are zero, and the problem is one of plane strain (see p. 60). The stresses may be found by means of a suitable stress function, or alternatively we may first solve the elastic equations to find the strains, and then deduce the stresses. We assume that the external surface of the cylinder (at infinity) is free from stress. Solutions have been given by many authors, especially Burgers (1939) and Koehler (1941). The three components of the displacements may be shown to be

$$\left. \begin{aligned} w_1 &= \{b/4\pi(1-\nu)\} \{2(1-\nu) \tan^{-1}(x_2/x_1) + x_1x_2/(x_1^2+x_2^2)\}, \\ w_2 &= -\{b/4\pi(1-\nu)\} \{(1-2\nu) \ln(x_1^2+x_2^2) - x_1^2/(x_1^2+x_2^2)\}, \\ w_3 &= 0, \end{aligned} \right\} \quad (30.6)$$

where  $\nu$  is Poisson's ratio. The displacement  $w_1$  is a continuous but not a single valued function; every time we describe a circuit round the dislocation line,  $w_1$  increases by  $b$ . This characteristic of all dislocation lines is a consequence of the body not being singly connected, and it was utilized in the formal definition on p. 245-7.

Differentiation of eqns. (30.6) gives us the components of the strain tensor, and the stresses are then obtained by Hooke's law. Since we are assuming the material to be elastically isotropic, there are only two independent constants in the stress tensor. From eqn. (11.26)

$$\left. \begin{aligned} X_{11} &= -B_e x_2(3x_1^2+x_2^2)/(x_1^2+x_2^2)^2, \\ X_{22} &= B_e x_2(x_1^2-x_2^2)/(x_1^2+x_2^2)^2, \\ X_{33} &= -2B_e \nu x_2/(x_1^2+x_2^2), \\ X_{12} &= X_{21} = B_e x_1(x_1^2-x_2^2)/(x_1^2+x_2^2)^2, \end{aligned} \right\} \quad (30.7)$$

where we have written  $B_e$  in place of  $\mu b/2\pi(1-\nu)$ . Alternatively, if we use cylindrical coordinates,  $r, \theta, x_3$ , the stress components are

$$\left. \begin{aligned} X_{rr} &= X_{\theta\theta} = -(B_e \sin \theta)/r, \\ X_{r\theta} &= (B_e \cos \theta)/r, \text{ and } X_{33} \text{ as above.} \end{aligned} \right\} \quad (30.8)$$

The stresses thus decrease with distance from the dislocation line, becoming zero at infinity in accordance with the assumed boundary conditions. For a dislocation in a real crystal, however, a more realistic boundary condition is that the external surface at some finite value of  $r$  shall be free from traction. This is achieved if we superimpose a second set of displacements which are single-valued and continuous, and which just cancel the stresses  $X_{rr}$  and  $X_{r\theta}$  at  $r = r_e$ , the external radius. Since these displacements are single-valued, they do not affect the dislocation. The necessary conditions are obtained if we superimpose the stresses

$$\left. \begin{aligned} X_{rr} &= (B_e/r_e^2)r \sin \theta, \\ X_{\theta\theta} &= (3B_e/r_e^2)r \sin \theta, \\ X_{r\theta} &= -(B_e/r_e^2)r \cos \theta, \end{aligned} \right\} \quad (30.9)$$

on to those of eqns. (30.8). The resultant stress  $X_{\theta\theta}$  given by these two sets of equations does not become zero at  $r = r_e$ , so that there is a tangential stress in the external surface caused by the edge dislocation in the material. Provided  $r_e$  is large, the stresses given by (30.9) are small in the region of the dislocation line, and may thus be neglected in comparison with those of (30.8).

The stresses on the internal surface of radius  $r_i$  (the radius of the central excluded region in the continuum approach) do not disappear when the stress fields of (30.8) and (30.9) are taken together. This corresponds to the true situation in a crystal, since this is not a real surface in the crystal, and the radius  $r_i$  is arbitrarily chosen. If  $r_i$  is very small, the displacements are large, and the stresses given by the continuum theory are much larger than the real stresses in a crystal; this is certainly true if we put  $r_i = b$ , so in general the excluded region must be a few interatomic distances wide. If we wish, we can free the internal surface from tractions by combining (30.8) with a stress function

$$\left. \begin{aligned} X_{rr} &= -X_{\theta\theta} = (B_e r_i^2 \sin \theta)/r^3, \\ X_{r\theta} &= -(B_e r_i^2 \cos \theta)/r^3. \end{aligned} \right\} \quad (30.10)$$

These stresses fall off rapidly with  $r$  and are only appreciable in the vicinity of  $r = r_i$ . Physically, we expect that the stresses at  $r = r_i$  will be intermediate between those given by (30.8) alone, and those zero values of  $X_{rr}$  and  $X_{r\theta}$  obtained by combining (30.8), and (30.10), but they should certainly be closer to the former assumption.

If we add together the three stress functions specified by eqns. (30.8), (30.9), and (30.10), we obtain a solution in which both internal and external surfaces are free from traction. Strictly, the constants in (30.9) and (30.10) have to be adjusted to give this result, but since the effect of (30.9) is negligible when  $r = r_i$ , and of (30.10) is negligible when  $r = r_e$ , the changes are only by factors of order  $(1 - (r_i^2/r_e^2))$ , which can be ignored.

Similarly, we may imagine a screw dislocation lying along the  $x_3$  axis of the cylinder, so that the Burgers vector is now parallel to  $x_3$ . The strain field is not now a plane deformation, but in fact the problem is simpler, since the displacement  $w_3$  is constant along the dislocation line, i.e. is independent of  $x_3$ , and this is the only component of the displacement, which is specified by

$$w_1 = w_2 = 0, \quad w_3 = (b/2\pi) \tan^{-1}(x_2/x_1), \quad (30.11)$$

and the corresponding stress components are

$$\left. \begin{aligned} X_{13} = X_{31} &= -B_s x_2 / (x_1^2 + x_2^2), \\ X_{23} = X_{32} &= B_s x_1 / (x_1^2 + x_2^2), \end{aligned} \right\} \quad (30.12)$$

where  $B_s = \mu b / 2\pi$ , and all other components of  $\mathbf{X}$  are zero.

The mathematical theory of screw dislocations is simpler than that of edge dislocations, because in an elastically isotropic medium only shear stresses and strains are involved. In cylindrical coordinates we have

$$w_3 = b\theta/2\pi, \quad X_{\theta 3} = \mu b/2\pi r, \quad (30.13)$$

and all other components of stress are zero. The stress field thus has radial symmetry, and depends only on  $r$ . This is related to the absence of a unique slip plane for the screw dislocation, so that the axes  $x_1$  and  $x_2$  are completely arbitrary. The multi-valued component of displacement  $w_3$  is independent of  $r$ .

We note that for a screw dislocation there is no traction across the outside surface of the cylinder of finite radius  $r_e$ . However, there is a couple on the end faces of a finite cylinder, so that to obtain correct boundary conditions in a finite specimen, we must superimpose stresses which cancel this couple and give single-valued displacements. The necessary stress field is

$$X_{\theta 3} = -\mu b r / \pi (r_e^2 + r_i^2), \quad (30.14)$$

where  $r_e$  is the external radius, as before, and  $r_i$  is the internal radius of the toroidal cylinder, and can be neglected in comparison with  $r_e$ . As in the case of edge dislocations, the additional stress field of (30.14) can be neglected reasonably close to the dislocation line, provided  $r_e$  is large.

The stress field of an arbitrary dislocation line will be more complex than that of the two simple types of Volterra dislocation which we have considered. More general formulae have been given by Peach and Koehler (1950) and other workers. The stresses all tend to zero as  $x_1$  or  $x_2$  tends to infinity, but the stress field extends throughout the material unless it is limited by other dislocations (see below).

Since we are interested in dislocations in real crystals, which are elastically anisotropic, we should really consider Volterra dislocations introduced into a continuum with anisotropic elastic properties. This theory is much more complex; it no longer follows that all the displacements produced by a screw dislocation must be parallel to the dislocation line, or that those produced by an edge dislocation must be normal to it. When all three components of displacement are non-zero, there is little advantage in treating edges and screws

separately, since the displacements due to an arbitrary element of line are not then separable in this way. However, the components of stress in the  $x_1x_2$  plane can be expressed in terms of the derivatives of  $w_1$  and  $w_2$  only, and the component of stress in the  $x_3$  direction in terms of derivatives of  $w_3$  only, provided the  $x_1x_2$  plane is a symmetry plane of the lattice. Several particular examples of dislocations in cubic structures have been solved in this way; the general theory and some important particular solutions were given by Eshelby *et al.* (1953).

The elastic theory of dislocations in an anisotropic medium is algebraically complex and cannot be described here in detail. Many numerical results are changed from the corresponding isotropic values by only  $\sim 20\text{--}50\%$ , which may not be significant in view of the approximations. However, anisotropic theory also leads to some results which are not predicted at all if isotropic theory is used; for example, in some metals straight dislocations in certain orientations are found to be unstable. For cubic structures the anisotropy constant (Table IV, p. 76) is usually a good measure of the deviations expected, and for highly anisotropic metals such as the alkali metals or  $\beta$ -brass these may be large. Although there are five independent elastic constants in a hexagonal material, there is, nevertheless, a large class of problems for which exact solutions may be obtained in hexagonal symmetry but not in cubic symmetry (Kröner, 1953; Eshelby, 1956a).

Eshelby *et al.* (1953) used a complex variable method to find the stress and strain fields of an infinite straight dislocation, and a slightly different method was developed by Stroh (1958). Except for certain symmetrical orientations (as noted above), the isotropic conditions  $w_1 = w_2 = 0$  for a pure screw dislocation and  $w_3 = 0$  for a pure edge dislocation do not apply, and a set of linear equations together with a sixth-order algebraic equation have to be solved. In general, analytical solutions may be obtained for dislocations parallel to symmetry axes and in planes normal to even-fold symmetry axes. Anisotropic solutions for the elastic problem of a general dislocation configuration have been obtained more recently; a general review is given by Bacon *et al.* (1978).

We may now calculate the energy of the dislocation in an otherwise perfect, elastically isotropic lattice. The energy is the sum of the elastic strain energy in the material, and the energy of the core of the dislocation, which cannot be treated by elastic theory. From eqn. (11.16), we see that the strain energy per unit volume at the point  $x_1x_2x_3$  is given by

$$dW_s/dv = B_e b/4\pi(x_1^2 + x_2^2) = B_e b/4\pi r^2,$$

for an edge dislocation, using the displacements and stresses of eqns. (30.6) and (30.7). Taking an annular volume element, and integrating from  $r = r_i$  to  $r = r_e$ , the internal and external radii of the cylinder,

$${}^L W_s = (B_e b/2) \ln(r_e/r_i), \quad (30.15)$$

as the strain energy per unit length of dislocation line.

More logically, we should use not the stresses of eqn. (30.8) alone, but those from (30.8), (30.9), and (30.10) together. This gives the slightly modified expression

$${}^L W_s = (B_e b/2) \{\ln(r_e/r_i) - 1\}. \quad (30.16)$$

The total energy of the dislocation line is obtained by adding the energy of the core of the dislocation to the elastic energy of (30.15) or (30.16). We thus have

$${}^LW_d = {}^LW_s + {}^LW_c \quad (30.17)$$

as the self-energy of the dislocation.  ${}^LW_c$ , the core energy, has been estimated in various ways, and shown to be between 0.5 and 2 eV per atom plane of the dislocation.

The same calculation applied to a screw dislocation gives for the strain energy per unit length of dislocation line

$${}^LW_s = (B_s b/2) \{\ln(r_e/r_i) - 1\}, \quad (30.18)$$

which differs from (30.16) only by a factor  $(1 - \nu)$ , which is between two-thirds and three-quarters for most metals.

Instead of integrating the elastic energy density, the energy of a dislocation line may alternatively be obtained directly as the work done on the surfaces of the cut (p. 257) whilst the displacement is increased from 0 to  $\mathbf{b}$ . Bullough and Foreman (1964) pointed out that this method requires inclusion of the work done by the tractions on the core surface (e.g. eqn. (30.10)), since otherwise inconsistent results are obtained from different orientations of the cut. However, the constant terms in eqns. (30.16) and (30.18) are normally unimportant, since the core radius is not defined in any case.

It follows from (30.16) and (30.18) that the strain energy of any straight dislocation line may be written

$${}^LW_s = \frac{1}{2} B b \ln(r_e/r_i), \quad (30.19)$$

where  $B$  varies between limits  $B_e$  and  $B_s$ . If the Burgers vector makes an angle  $\varphi$  with the line, it may be regarded as coincident screw and edge dislocations of vectors  $b \cos \varphi$  and  $b \sin \varphi$  respectively, so that

$$B = B_e(1 - \nu \cos^2 \varphi). \quad (30.20)$$

Also since the shear stress on the slip plane is  $B_e/r$  and  $B_s/r$  for edges and screw respectively (eqns. (30.8) and (30.13)) it also follows that

$$\tau_s = B/r \quad (30.21)$$

is the effective shear stress on the slip plane produced by any straight dislocation.

In an anisotropic medium eqn. (30.19) is modified (Foreman, 1955; Stroh, 1958), and becomes

$${}^LW_s = (K b^2/4\pi) \ln(r_e/r_i), \quad (30.22)$$

where  $K$  is a function of the elastic constants and of the orientations of both the dislocation line and its Burgers vector. For symmetrical orientations,  $K$  may be expressed in analytical form; and in the general case it may be written as a Fourier series

$$K = \sum_{n=0}^{\infty} (\alpha_n \cos n\varphi + \beta_n \sin n\varphi). \quad (30.23)$$

Equations (30.19) and (30.22) show that in an infinite crystal the strain energy of a single dislocation diverges logarithmically to infinity. In a real crystal there is a finite strain energy which is about 10 eV per atom plane if  $r_e$  is taken as  $10^{-2}$  m and we assume typical values for elastic constants and lattice parameters. Actually the effective radius of the elastic field of each dislocation in a real crystal is limited by the presence of other dislocations, and the integration should probably be cut off at a separation  $r_e = r_d$ , where  $r_d$  is the mean separation of dislocations. A typical value of  $r_d$  is 1  $\mu\text{m}$ , and the strain energy of the dislocation is then rather less than half the value quoted above, but is still several electron volts per atom plane.

It was formerly assumed that a random distribution of dislocations would correspond to a cut-off radius  $r_d$ , but Wilkens (1967, 1968) showed that the mean interaction energy of any two dislocations vanishes in a random arrangement, so that the energy of each dislocation diverges logarithmically with a radius of the order of the specimen size. This implies that a random arrangement is prevented by the dislocation interactions and the real arrangement will probably be such that the effective  $r_e$  is not appreciably larger than  $r_d$ .

Equation (30.17) represents the internal or self-energy of the dislocation. In order to calculate the free energy, it is necessary to include the effects of entropy terms, and as with the defects previously considered, these may be separated into the effects of the dislocation on the vibrational spectrum of the crystal (thermal entropy), and the configurational entropy of the dislocation itself. Cottrell (1953) has shown that both these effects are negligible in comparison with strain energy, even if the dislocation is assumed to be completely flexible on an atomic scale. We thus conclude that the expression for the self-energy of a dislocation also gives its free energy to a close approximation. Since this energy is positive, and of magnitude very much larger than  $kT$ , the dislocation cannot exist as a thermodynamically stable defect. This is in contrast to the point defects considered in Chapter 5. It follows from the form of eqns. (30.16) and (30.18) that to a good approximation the free energy of any dislocation may be assumed to be proportional to  $b^2$ . There is a small dependence on the orientation of the line, which may be neglected since the factor  $B$  in the energy equation in nearly isotropic crystals only varies between about  $\frac{2}{3}$  and 1. The dependence on crystal anisotropy, is important, however, in some special cases. The dependence of the dislocation energy on  $b^2$  is important in the theory of dislocation reactions.

When two dislocations are present in a crystal, the energy of the stress field can be formally represented as the sum of the self-energies of the two dislocations plus a term representing the energy of interaction. If the two dislocations have equal and opposite Burgers vectors, the strains produced will cancel at distances large compared with their separation,<sup>†</sup> and the energy density thus tends to zero except in the region around the dislocations. Two equal and opposite dislocations have finite energy even in an infinite crystal, and the same conclusion applies to a closed dislocation loop.

The problem of two dislocations can be treated by the principle of superposition (p. 64). The integration of the part of the strain energy which represents the interaction of the dislocations is, however, difficult, and may be avoided by means of a method due to Cottrell

<sup>†</sup> More correctly, the displacements will fall off as  $r^{-2}$ , and the strain energy integral is thus convergent.

(1948, 1953). We imagine one dislocation to be already present in the crystal, and consider the work done when a second dislocation is introduced. This work may be divided into work done against the elastic resistance of the material, which represents the self-energy of the second dislocation, and the work done on or by the stress field of the first dislocation, which represents the interaction energy.

Consider first a positive edge dislocation lying along the  $x_3$  axis with Burgers vector along  $Ox_1$  and the extra plane of atoms  $Ox_2x_3$ . Let a second, parallel dislocation be brought into the crystal so that it lies along a line distant  $r$  from the  $x_3$  axis and having equation  $x_1 = r \cos \theta$ ,  $x_2 = r \sin \theta$ . This dislocation may be formed by making a cut parallel to the  $Ox_2x_3$  plane and inserting an extra plane of atoms or vacancies, the boundaries of the cut being displaced by  $b$  in the  $x_1$  direction. Taking a unit length of this dislocation parallel to  $Ox_3$ , the forces due to the first dislocation acting across an element of area  $dx_2$  of the  $Ox_2x_3$  plane are  $X_{11} dx_2$ , and when a positive dislocation is formed, the work done by the forces acting on this area is  $X_{11} b dx_2$ . The total gain in energy is thus obtained by integrating the negative of this expression between limits  $x_2 = r \sin \theta$  and  $x_2 = r_e \sin \theta$ . After substituting for  $X_{11}$  from eqn. (30.7), the interaction energy is thus obtained as

$${}^L W_i = B_e b [\ln r_e - \ln r + \cos^2 \theta], \quad (30.24)$$

where  $(x_1/r_e \sin \theta)^2$  has been neglected since  $r_e \gg x_1$ .

The force between the dislocations may now be obtained by differentiating this expression. We find

$$\left. \begin{aligned} f_r &= \partial {}^L W_i / \partial r = B_e b / r, \\ f_\theta &= -(1/r) \partial {}^L W_i / \partial \theta = (B_e b / r) \sin 2\theta, \end{aligned} \right\} \quad (30.25)$$

as the radial and tangential forces per unit length of either dislocation. The radial force is positive (repulsive) for dislocations of the same sign, but the existence of the  $f_\theta$  component shows that it is not central. For opposite dislocations, the radial force is attractive.

Since the dislocations are edge types, they can glide only in a particular plane. The component of force providing glide motion is

$$\left. \begin{aligned} f_1 &= f_r \cos \theta - f_\theta \sin \theta \\ &= (B_e b / r) \cos \theta \cos 2\theta \\ &= (B_e b / 4y) \sin 4\theta, \end{aligned} \right\} \quad (30.26)$$

where  $y$  is the normal separation between the glide planes. For dislocations of like sign, the force  $f_1$  as a function of  $\cot \theta$  is shown in Fig. 7.14. We see that the dislocations attract each other (i.e. the force acts so as to increase  $\theta$ ) for  $\theta > \pi/4$  and repel one another for  $\theta < \pi/4$ . The force is zero at  $\theta = \pi/2$  and  $\theta = \pi/4$ , but whereas the former equilibrium is stable, the latter is unstable, a small displacement to lower or higher angles leading to repulsive or attractive forces respectively. A set of edge dislocations of like sign on parallel slip planes are thus in stable equilibrium when they line up in the direction perpendicular to the slip planes; this is further discussed in Chapter 8. It also follows that two dislocations of unlike sign on different parallel slip planes are in stable equilibrium when  $\theta = \pi/4$ .

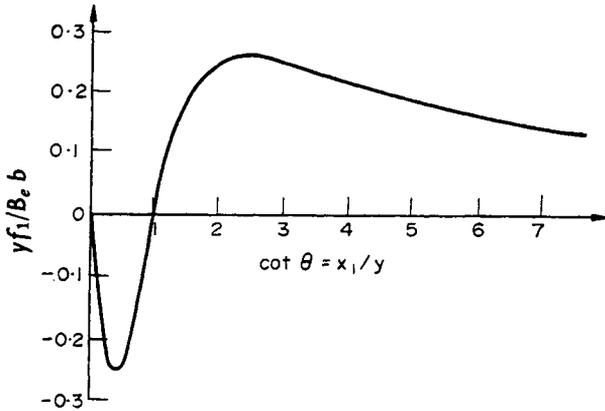


FIG. 7.14. Component of force acting in the glide plane between two like edge dislocations moving in glide planes with a separation of  $y$  in the  $x_2$  direction.

We can perform an identical calculation for two parallel screw dislocations, to obtain

$${}^LW_i = B_s b [\ln r_e - \ln r], \tag{30.27}$$

and the corresponding forces are

$$f_r = B_s b / r, \quad f_\theta = 0. \tag{30.28}$$

The radial forces between two parallel dislocations with parallel Burgers vectors thus vary only between the limits represented by the factor  $B_e$  or  $B_s$ . For the edge dislocations, the tangential force  $f_\theta$  disappears when the dislocations have a common slip plane. This condition is automatically satisfied for two parallel screw dislocations, and there is thus no tangential component of force.

The total strain energy for two parallel dislocations of opposite sign is obtained by adding the expression for the self-energies (eqns. (30.15)–(30.18)) to the interaction energy (eqns. (30.24) and (30.27)). Thus for unit length of an edge dislocation pair, or dipole,

$${}^LW = B_e b [\ln (r/r_i) - \cos^2 \theta] + 2 {}^LW_c, \tag{30.29}$$

and there is a similar expression for a screw dislocation dipole. Since  $r_e$  does not appear in eqn. (30.29), the energy per unit length of such a dipole is finite even in an infinite crystal, as anticipated above. In this approximation, the binding energy of the edge dipole (with respect to dissociation by glide) is obtained by putting  $\theta = \pi/4$  in eqn. (30.24), and is thus approximately  $B_e b \ln(r_e/y)$ , where  $y$  is the separation of the glide planes. This is a large energy, comparable with the self-energy when  $y \ll 100b$ .

Edge dipoles and multipoles (see p. 315) are often observed as a principal feature of the dislocation structure in the early stages of deformation of single crystals. Screw dipoles are observed much less frequently, presumably because the component dislocations can often annihilate each other by cross-slip. A closed loop may be regarded as composed of infinitesimal dipole elements, and so its energy will also be finite. In first approximation,

the energy is  $(Bb/2) \ln(r/r_i)$  per unit length of the loop, so that the energy of a loop of radius  $r$  is

$$W_d = \frac{1}{2} \mu b^2 r \ln(r/r_i). \quad (30.30)$$

A more accurate expression is given below.

The above expressions were derived for parallel dislocations with equal, or equal and opposite, Burgers vectors. More general relations have been given by Bilby (1950) and Nabarro (1952). For two parallel screw dislocations with unequal Burgers vectors we have obviously merely to replace  $b^2$  by  $b_1 b_2$ , and the same applies to edge dislocations with collinear Burgers vectors. Two parallel edge dislocations with non-parallel Burgers vectors  $\mathbf{b}_1$  and  $\mathbf{b}_2$  give an interaction energy in which the term  $b^2 \ln r$  in eqn. (30.24) is replaced by  $\mathbf{b}_1 \cdot \mathbf{b}_2 \ln r$ , and the term  $b^2 \cos^2 \theta$  is replaced by  $(\mathbf{b}_1 \cdot \mathbf{r})(\mathbf{b}_2 \cdot \mathbf{r})/r^2$ . The latter term equals  $b_1 b_2 \cos^2 \theta$  when the two Burgers vectors are parallel, and  $\frac{1}{2} b_1 b_2 \sin 2\theta$  when they are perpendicular to each other.

More generally, two parallel dislocation lines lying along a direction  $\mathbf{i}$  and having edge and screw elements  $\mathbf{b}_{e1} \mathbf{b}_{e2} \mathbf{b}_{s1} \mathbf{b}_{s2}$  (see p. 242) have an energy of interaction given by

$${}^L W_i = \frac{\mu}{2\pi} \left[ b_{s1} b_{s2} \ln \frac{r_e}{r} + \frac{1}{1-\nu} \left\{ \mathbf{b}_{e1} \cdot \mathbf{b}_{e2} \ln \frac{r_e}{r} + \frac{(\mathbf{b}_{e1} \cdot \mathbf{r})(\mathbf{b}_{e2} \cdot \mathbf{r})}{r^2} \right\} \right].$$

The force between the dislocations is obtained by differentiating this expression.

Parallel dislocations exert a force on each other which has a constant magnitude per unit length. Perpendicular dislocation lines, in contrast, only interact strongly in the region where they approach closest to each other. The relevant formulae have been summarized by Nabarro (1952). There is no interaction between perpendicular dislocations unless the Burgers vector of each is parallel to the line of the other; when this condition is satisfied there is a finite interaction which does not lead to glide. The more general interaction between two nonparallel dislocations on intersecting glide planes is very complex; it is important in some theories of work hardening (Nabarro *et al.*, 1964). Dislocations which pass through each other produce jogs, as described on p. 252.

The displacement field of a general dislocation loop in isotropic approximation was first derived by Burgers (1939) as an integral over the cut surface of the loop; it may alternatively be found by a Green's function method (Seeger, 1955a, b; de Wit, 1960). In either case, the algebra is unwieldy, although numerous particular solutions have been found. For a circular loop in the glide plane, Kröner (1958) found the stress field in the form of complete elliptical integrals, and the corresponding strain energy is

$$W_d = \frac{1}{2} \pi (2-\nu) B_e b r \{ \ln(4r/r_i) - 1 \}. \quad (30.31)$$

This is insignificantly different from the approximate expression (30.30). The stress field of a circular pure prismatic loop is obtained as definite integrals of Bessel functions (Kroupa, 1960; Bullough and Newman, 1960), and the energy is

$$W_d = \pi B_e b r \{ \ln(8r/r_0) - 1 \}. \quad (30.32)$$

A general method for the calculation of self-energies and interaction energies in which any dislocation line is constructed from angular dislocations was developed by Yoffe (1960), and a similar approach in which the dislocation is treated as piecewise straight is due to Jøssang, *et al.* (1965), de Wit (1967), and others. The energies of some complex loop shapes have been determined analytically by these methods; these calculations were reviewed by Kroupa (1966).

An anisotropic calculation of the elastic field and associated energy of a general dislocation configuration has only recently become possible. Indenbom and Orlov (1967) solved the elastic problems of a finite dislocation segment by generalizing approaches due to Lothe (1967) and Brown (1967). Their solution utilizes the derivatives with respect to orientation of the distortions produced by an infinite straight dislocation. Analytical solutions which do not require solution of a set of linear or algebraic equations have been derived by Willis (1970), also by means of a Green's function method. Willis gives for the infinite straight dislocation explicit (though complex) expressions which are easier to apply than the formulations of Eshelby *et al.* (1953) and Stroh (1958), and he also derives equations for the distortions produced by a dislocation segment and by any planar dislocation loop. These expressions do not require the analytical solutions for infinite, straight dislocations, and can be used directly for the calculation of the stress and strain fields or energies of any dislocation configuration. A further simplification of the computing procedure has been suggested by Barnett and Swanger (1971), Barnett *et al.* (1972), and Barnett (1972), who extended Brown's method and gave formulae for the line energy  ${}^LW_s$  and its first and second derivatives with respect to orientation. Their procedure avoids the necessity for solving the sextic algebraic equation by using a numerical integration method based on a Fourier transform. Alternative numerical procedures are given by Malen (1970) and Malen and Lothe (1970). These important techniques have been applied to the elliptical loop (Willis, 1970), the rhombus-shaped loop (Bacon *et al.*, 1970) and other cases; for details see the review by Bacon *et al.* (1978).

We next consider briefly the concept of the line tension of a dislocation which was introduced by Mott and Nabarro (1948) who used the taut-string analogy mentioned above and defined the line tension as the increase in energy for a unit increase in line length. The line tension  $T$  thus has the dimensions of force and for a straight dislocation in isotropic approximation is given from eqn. (30.19) as

$$T = {}^LW_s = \frac{1}{2}Bb \ln(r_e/r_i). \quad (30.33)$$

However, the Mott and Nabarro definition gives a line tension which depends on the orientation and shape of the dislocation; obviously a slightly different expression is obtained by differentiating (30.31) to obtain the line tension in a circular loop. De Wit and Koehler (1959) first suggested that any change in the orientation of the dislocation line should be included in the line tension, which would thus be defined as

$$T = {}^LW_s + (\partial^2 {}^LW_s / \partial \varphi^2), \quad (30.34)$$

where  $\varphi$  is the angle between  $\mathbf{b}$  and the direction of the line element. A more detailed study of this problem was made by Brown (1964), who introduced the concept of self-stress, i.e.

the stress on an element of dislocation due to the dislocation itself. This stress varies along the dislocation with the local orientation, and also depends on the shape of the whole line; it is expressed as a rather complex line integral. If the curvature is everywhere small, the self-stress is well-approximated by a simple tension, as assumed above, but if this condition is not satisfied it is necessary to integrate numerically.

Further calculations of the self-stress have been made by Jøssang *et al.* (1965), Jossang (1968), and other workers who treated a bowed dislocation as composed of piecewise straight segments, and by Brailsford (1965) who approximated a curved dislocation by a sequence of infinitesimal kinks. These results are important in certain problems in dislocation theory, e.g. the use of measurements on extended nodes to estimate stacking fault energies, but for many purposes it is adequate to use the rough approximation  $T = \mu b^2$  which follows from (30.33) if  $r_e \sim 10^4 b$  and  $r_i \sim b$  (Friedel, 1964).

Line tension may be used to estimate the equilibrium shape of a dislocation with fixed ends in an external or internal stress field. An element  $dd$  which subtends an angle  $d\psi$  at its centre of curvature is in equilibrium under the force per unit length  $f$  from the stress field and its own line tension  $T$  if

$$T d\psi = f dd, \quad (30.35)$$

and hence the curvature is

$$\frac{1}{r} = d\psi/dd = f/T. \quad (30.36)$$

For a Frank-Read source of length  $d$ , the critical condition is when the semicircular configuration is reached, so that  $r = d/2$ . If the resolved shear stress in the direction of  $b$  is  $\tau$  the source will operate if

$$r \geq 2T/db. \quad (30.37)$$

More accurately, the dislocation will not adopt a semicircular configuration, even in isotropic approximation, and the equilibrium shape must be determined by the condition that the total force (from imposed stresses, internal stresses, and self-stress) on each element of the line must be zero (Brown, 1964; Bacon, 1967; Foreman, 1967). Moreover, when anisotropic theory is used a new result appears, namely that it is possible for the line tension of a straight dislocation line to be negative (de Wit and Koehler, 1959). This means that the straight line is unstable, and if its mean direction is fixed it will adopt some kind of zigzag configuration. The conditions leading to negative line tension in several metals have been calculated by Head (1967) and Barnett *et al.* (1972) and are further discussed in Section 33.

In the above discussion we have ignored any internal or external surfaces, except insofar as the latter determine  $r_e$ . However, the stress field of a dislocation near an interface or a free surface is modified by the requirement that the shear and normal components must be continuous across the boundary so that there is an interaction of the dislocation with the interface. Solutions to this elastic problem can often be obtained by the method of images which is exactly analogous to that used in electrostatics. Imaginary sources of stress, e.g. other dislocations, are introduced so that the superimposed stress fields can be made to

satisfy the boundary conditions. The stress field of an image source appears as an external stress field to the real dislocation, and the resultant force on the dislocation is called an image force.

For a screw dislocation in a medium of shear modulus  $\mu_1$  near a planar boundary with a medium of shear modulus  $\mu_2$ , the image dislocation has a Burgers vector

$$\mathbf{b}_I = \{(\mu_2 - \mu_1)/(\mu_2 + \mu_1)\}\mathbf{b} \quad (30.38)$$

(Head, 1953), and the dislocation experiences a force normal to the boundary of magnitude

$$f = B_s b_I / 2r, \quad (30.39)$$

where  $r$  is the distance of the dislocation from the boundary. The dislocation is attracted to the boundary if  $\mu_2 < \mu_1$  and is repelled if  $\mu_2 > \mu_1$ ; in particular it is attracted to a free surface ( $\mu_2 = 0$ ). Similar calculations have been made for an edge dislocation (Head, 1953), a dislocation inclined to the surface (Lothe, 1967), and a general dislocation loop (Groves and Bacon, 1970). Image forces are obviously of great importance in considering dislocation configurations observed by transmission electron-microscopy of thin foils.

In addition to the forces which they exert on each other, dislocations will also interact with the stress fields of other types of lattice defect. Of particular interest is the interaction between a dislocation and a solute atom, which may be dissolved substitutionally or interstitially in the lattice, or between a dislocation and a point defect. We have already seen that an edge dislocation produces an expanded region on one side of its slip plane, and a compressed region on the other. If the solute atom is larger than the solvent atom, or if it is dissolved interstitially, we expect it to be repelled from the compressed side and attracted into the expanded side. Conversely, a small solute atom or a lattice vacancy will be attracted to the compressed region; the vacancy may disappear completely at a jog, as discussed earlier, or may be merely bound to the dislocation line in the same way as a solute atom.

A simple expression for the energy of interaction of a dislocation line and a solute atom was first given by Cottrell (1948). He used the misfitting sphere model of the solute atom which we have already discussed in Section 25. The self-energy of the solute atom in this model is equal to the work done against the elastic resistance of the crystal in forcing a spherical atom into a spherical hole of different size. In calculating this energy previously, we made the implicit assumption that the crystal is free from internal stress fields. If an internal stress field is present, however, additional work will be done on or by the forces of this field, and this work may be termed the energy of interaction. The method is clearly identical with that used above for finding the energy of interaction of two dislocation lines.

Suppose the change in volume is  $\Delta v$  when a single solute atom is introduced. Then if the stress field is assumed uniform over the region occupied by the atom, and not to change during the expansion, the work done against the stress field is  $p \Delta v$ , where  $p$  is the hydrostatic pressure in the region of the atom. We thus have

$$W_i = p \Delta v = K \Delta \Delta v, \quad (30.40)$$

where  $K$  is the bulk modulus of elasticity, and  $\Delta$  is the cubic dilatation near the atom due to the stress field of the dislocation. Eshelby (1954, 1956a) has shown that in this expression  $\Delta v$  is the change in volume of the whole crystal (including "image" effects) when the point defect is introduced. Using the result that the hydrostatic pressure in a stress field  $\mathbf{X}$  is  $p = -X_{ii}/3$  we have from eqn. (30.8)

$$W_i = \frac{\mu b}{3\pi} \frac{1+\nu}{1-\nu} \frac{\sin \theta}{r} \Delta v. \quad (30.41)$$

For a solute atom which expands the lattice, this is positive when  $0 < \theta < \pi$ , i.e. when the atom is above the slip plane, and negative when the atom is below the slip plane. The atom is thus expelled from the compressed region and attracted into the expanded region.

If the model of a solute atom as a compressible sphere in a misfitting hole is used, the interaction energy may also be written (Bilby, 1950) as

$$W_i = \frac{\mu b}{\pi} \frac{\sin \theta}{r} \Delta v^\infty. \quad (30.42)$$

where  $\Delta v^\infty$  (see p. 203) is the change in volume of a region enclosed by a surface immediately surrounding the solute atom. Equation (25.8) gives the relation between  $\Delta v^\infty$  and the difference  $\Delta v_{AB}$  in atomic volumes of the two components. In principle,  $W_i$  may be obtained from an estimate of  $\Delta v_{AB}$ , but in practice it is preferable to use measurements of the change of lattice parameter with composition to obtain the mean strain  $\epsilon = d \ln a/dx$  (see p. 185) and then to substitute  $\Delta v = 3v\epsilon$  directly into (30.41).

The existence of an interaction energy implies that there is an effective force between an edge dislocation and a solute atom (or a vacancy) at any point. The force components  $f_r = -\partial W_i/\partial r$  and  $f_\theta = -(1/r)(\partial W_i/\partial \theta)$  may be evaluated from (30.41) or (30.42).

Clearly this type of calculation suffers from all the defects mentioned in Section 25, but it is likely to be qualitatively correct. Even if the elastic model for the solute atom is accepted, however, the calculation is only applicable in the region where the internal stress field of the dislocation is given by linear elastic theory. The expression should thus not be applied directly to solute atoms which are very close to the dislocation axis. When  $\Delta v$  is positive, it is clear that the strongest binding (lowest value of  $W_i$ ) is obtained at  $\theta = 3\pi/2$  and at some finite distance  $r$  which will be about one interatomic distance. The change in the stress field of the dislocation due to the presence of solute atoms near the dislocation is also not considered in this simple form of calculation.

Cottrell supposed the strain field of the solute atom to be a pure dilatation, so that it interacts only with hydrostatic stress. When elastic anisotropy is taken into account, this conclusion is no longer valid, but for symmetrical solute atoms (or point defects) the main energy will still arise from the dilatational component of the dislocation field. An interesting result which follows from the different geometries is that a screw dislocation has a dilatational field in b.c.c. structures, but not in f.c.c. or h.c.p. structures (Chang, 1962; Chou, 1965). Thus even in the linear elastic approximation, screw dislocations interact with centres of dilatation in b.c.c. crystals. However, a more important effect may be the presence of solutes in sites of lower than cubic symmetry.

Point defects which do not possess the spherical (or cubical) symmetry required for Cottrell's theory have a long-range elastic interaction with all dislocations. Tetragonal defects are of especial interest since they represent interstitial solutes in b.c.c. crystals and various other defects in f.c.c., ionic or diamond cubic structures. In many b.c.c. metals, interstitial solutes such as carbon, nitrogen, or oxygen are believed to occupy octahedral sites, that is equivalent sites at the centres of {100} faces and  $\langle 100 \rangle$  edges of the unit cube. The resultant shear strains can interact with the shear stress field of a screw dislocation to give an appreciable energy (Crussard, 1950; Cocharadt *et al.*, 1955). If the defect is specified by principal strains  $e_{11} = e_1$  and  $e_{22} = e_{33} = e_2$ , the interaction energy depends on the shear strain  $(e_1 - e_2)$  and according to Cocharadt *et al.* it is given by

$$W_i = 8^{\frac{1}{2}} B_s (e_1 - e_2) v \sin \theta / r, \quad (30.43)$$

where  $\theta$  is now the smallest angle between the radius vector from the dislocation to the solute atom and a  $\langle 110 \rangle$  direction normal to the dislocation line. The volume  $v$  with which the strains  $e_1$  and  $e_2$  are associated is effectively an ellipsoid of revolution, but is of ill-defined size. However, if the values of  $e_1$  and  $e_2$  are derived from experimental measurements of the lattice parameters of aligned defects in dilute solution (as is possible for carbon or nitrogen martensites), the choice of  $v$  is immaterial. The lattice parameter measurements are extrapolated to one defect per volume  $v$  and the product  $(e_1 - e_2)v$  is then independent of  $v$ . Some confusion has arisen because Cocharadt *et al.* chose to regard the volume  $v$  as the cubic cell volume  $a^3$ . This implies rather large values for the principal strains, and Hirth and Cohen (1969a, b) argued that a more reasonable choice of  $v$  is  $\frac{1}{2}a^3$ , which effectively doubles these large strains. These authors therefore argued that the lattice parameter extrapolation procedure is invalid and that relative displacements of nearest neighbours obtained by an atomistic calculations should be used in (30.43). Schoeck (1969) and Bacon (1969) have shown that this is incorrect and that (30.43) is exact within the limitations of continuum theory.

The interaction of a tetragonal defect with an edge dislocation is given by

$$W_i = \left\{ \frac{1}{3}(e_1 + 2e_2)(1 + \nu + 2 \cos^2 \theta) - e_2 \cos 2\theta + e_1 \nu \right\} (B_e v / r) \sin \theta. \quad (30.44)$$

The interactions with edge and screw dislocations are of comparable magnitude.

The above interactions all involve energies falling off as  $1/r$  and may be described as long-range. Shorter-range interactions, decreasing as  $1/r^2$ , arise from a model in which the solute atom is treated as an elastic inhomogeneity, i.e. as a finite volume of different elastic properties. For example, an atom of slightly different shear modulus has an interaction energy with a screw dislocation (Saxl, 1964),

$$W_i = e_\mu v {}^v W = (\mu / 8\pi^2) e_\mu (b/r)^2. \quad (30.45)$$

where  ${}^v W$  is the strain energy density of the dislocation and  $e_\mu$  is the fractional rate of change of  $\mu$  with atomic fraction of solute. A similar equation for an edge dislocation involves the rate of change of both bulk and shear moduli.

Finally, there is a second-order volume expansion around a screw dislocation of magnitude  $Cb^2/4\pi^2r^2$ , where  $0.3 \ll C \ll 1$  (Stehle and Seeger, 1956). This gives a short-range interaction

$$W_i = \left\{ \frac{C(1+\nu)}{2\pi^2(1-2\nu)} \right\} \mu ve \left( \frac{b}{r} \right)^2. \quad (30.46)$$

The change of volume arises from anharmonic vibrations, and Lomer (1957) pointed out that  $C$  may be derived from the Grüneisen relation between thermal expansion and specific heat.

A full discussion of the interaction energy between a line dislocation and a point defect should include electrical and chemical interactions as well as elastic interactions. Generally, the electrical interaction is small for metals in comparison with the elastic interaction, though this is not necessarily true for polar crystals. The chemical interaction arises mainly when the dislocation is extended into a ribbon of stacking fault, since a lowering of energy can result from a change of composition in the stacking fault region. This effect was first discussed by Suzuki (1952).

The maximum value of the interaction energy (the binding energy) of a solute atom and a dislocation may be estimated theoretically or from experimental measurements. The calculations of Cocharadt *et al.* suggest a binding energy of about 0.75 eV for a carbon atom in iron for both screw and edge dislocations. This is almost certainly too high, and various other estimates fall in the region 0.3–0.5 eV, but the results are significant in showing that for b.c.c. metals with interstitial solutes, the binding energies are comparable for all types of dislocations. For f.c.c. metals, where solutions are usually substitutional, this is not true, and edge dislocations are bound more strongly than screw dislocations, at least if chemical binding is ignored for the present. Typical binding energies for substitutional solutes are probably about 0.02 eV, and this should also be about the order of magnitude of the interaction between a dislocation and a vacancy.

Provided the necessary diffusion of solute atoms is possible, we may expect the interactions to lead to a segregation of solute atoms to the vicinity of a dislocation. The equilibrium state of a crystal containing both dislocations and impurities will thus be one in which each dislocation is surrounded by a cloud of impurity atoms; this has been called by Cottrell (1948) a dislocation "atmosphere". Under suitable conditions, the equilibrium solute distribution around a dislocation may be given by a Maxwellian formula of the type

$$c = c_0 \exp(W_i/kT), \quad (30.47)$$

where  $c_0$  is the mean solute concentration. This equation will be valid so long as the change in concentration which it predicts at any point is small, i.e. the atmosphere is "dilute". The necessary conditions for this are that  $c_0$  is small and  $W_i \ll kT$ . If these conditions are not satisfied, the atmosphere will condense, probably into a row of solute atoms parallel to the dislocation and situated at the points of maximum binding. A rough condition for the condensation of the atmosphere is that the dilatations due to the solute atoms just cancel those due to the dislocation, so that the hydrostatic stress is completely relaxed. This happens approximately when  $(c - c_0)\Delta v$  is equal to the maximum value of the dilatation near the dislocation centre. In the case of interstitial solutes in b.c.c. metals, the interaction

is so strong that  $W_i \gg kT$  at room temperature, thus leading to condensed or saturated atmospheres. As we have already seen, this strong interaction does not arise primarily from purely hydrostatic stresses.

Even for a saturated atmosphere, there should be only one interstitial atom per atom plane of the dislocation within a radial distance of a few Ångstrom from the dislocation core. Clearly, this distribution is not described by eqn. (30.47); Louat (1956) has suggested instead the use of a distribution law of the Fermi type. We should also note here that there is much evidence that at low-temperatures dislocation lines in metals like iron acquire a very large number of interstitial atoms per atom plane of dislocation. This is, in fact, a precipitation phenomenon, the dislocation line acting as a centre on which the new phase precipitates. The theory of the segregation of solute atoms to dislocations is obviously closely related to that of precipitation proper in certain circumstances, and we shall discuss this in more detail in Part II, Chapter 16.

### 31. FORMATION AND PROPERTIES OF CRYSTAL DISLOCATIONS

In the last section we obtained explicit expressions for the stress field of a dislocation by treating the solid as an elastic continuum. This procedure is only valid if the region around the dislocation line is excluded from consideration. The stress field at the centre of a dislocation is infinite in an elastic continuum, but in a crystalline material, the shear stress between the planes which glide over each other must be a periodic function of  $b$ . In the model of a dislocation used by Peierls (1940) and extended by Nabarro (1947), the half crystals on

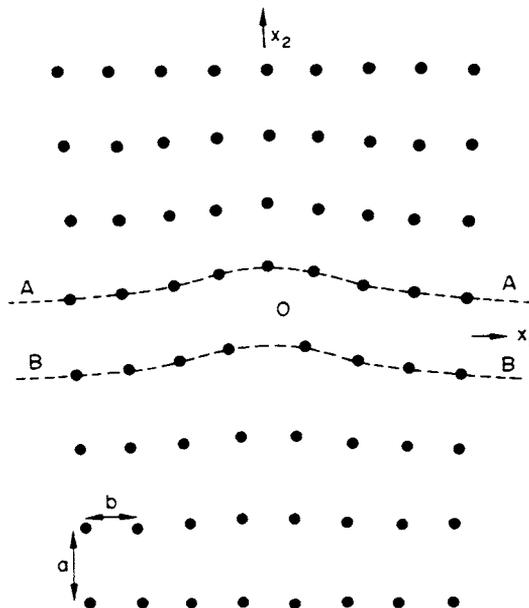


FIG. 7.15. The Peierls-Nabarro model of a dislocation (after Nabarro 1952).

either side of the glide plane are treated as elastic continua, but the shear stress between two atomic planes is assumed to be a sinusoidal function of the relative displacement of the atoms.

Figure 7.15 shows a section through a straight-edge dislocation lying along the  $x_3$  axis. The two planes  $A$  and  $B$  are curved by the dislocation, but in the model used there is no normal stress  $X_{22}$ . If the interatomic distances in the  $x_1$  and  $x_2$  directions are  $b$  and  $a$  respectively, the Burgers vector has magnitude  $b$ ; for a simple cubic crystal as used in most of the calculations,  $a = b$ . Let the distance along the  $x_1$  direction between neighbouring atoms in the planes  $A$  and  $B$  be written

$$\Phi_1(x_1) = \begin{cases} 2w_1(x_1) + \frac{1}{2}b & (x_1 > 0), \\ 2w_1(x_1) - \frac{1}{2}b & (x_1 < 0), \end{cases} \quad (31.1)$$

where  $w_1, -w_1$  are the components in the  $x_1$  direction of the displacements of the atoms in  $A$  and  $B$  respectively, measured from an initial configuration in which the atoms in the two planes are all out of registry by  $\frac{1}{2}b$  but the planes themselves are undistorted. Any components of displacement  $w_2$  are assumed to be equal in  $A$  and  $B$ . The choice of origin for  $w$  means that the displacements tend to zero near the centre of the dislocation line if the core has the symmetrical configuration shown in Fig. 7.15. There is also a boundary condition  $w_1(\infty) = -w_1(-\infty) = -b/4$  which ensures that the atoms are in registry ( $\Phi_1 = 0$ ) far from the centre.

The width of the dislocation is the region over which the disregistry is appreciable, and is determined by the balance between two opposing forces. The atoms in  $A$  are compressed in the  $x_1$  direction, and try to reduce their energy by spreading this compression uniformly along the plane; this force is calculated by elastic theory. The opposing force is due to the extended atoms in  $B$  which attempt to pull those in  $A$  closer together, and thus reduce the dislocation width. Peierls and Nabarro assumed that the shear stress  $X_{12}$  between the planes can be expressed as a sinusoidal function of displacement of the form

$$X_{12} = (-\mu b/2\pi a) \sin(4\pi w/b), \quad (31.2)$$

where the constant is fixed by requiring Hooke's law to be valid in the limit of small shear.

The elastic shear stress in  $A$  due to the displacements  $w_1(x_1)$  may be calculated directly or, alternatively, by noting (Eshelby, 1949) that the model is equivalent to a continuous distribution of infinitesimal elastic edge dislocations on the slip plane with Burgers vector between  $x_1$  and  $x_1 + dx_1$  of  $-2 dw_1(x_1)$ . By considering the shear stress at  $x_1$  due to the infinitesimal dislocations lying between  $x_1 = x', x' + dx'$  (eqn. 30.7) and integrating,

$$X_{12}(x_1, 0) = \frac{\mu}{\pi(1-\nu)} \int_{-\infty}^{\infty} \left\{ \frac{[dw_1(x_1)/dx_1]_{x_1=x'}}{x_1 - x'} \right\} dx'. \quad (31.3)$$

Equations (31.2) and (31.3) give the integral equation for the displacements  $w_1$  and the solution which satisfies the boundary conditions at infinity is

$$w_1 = -(b/2\pi) \tan^{-1}\{x_1 + \alpha b/\zeta\}, \quad (31.4)$$

where

$$\zeta = a/2(1-\nu) \quad (31.5)$$

and  $0 \leq \alpha \leq 1$ . The disregistry  $\Phi_1$  falls from its maximum value to one-half of this value in a distance  $\zeta$ , so that  $2\zeta$  is a measure of the width of the core. The parameter  $\alpha$  specifies the symmetry of the core, and is such that  $\alpha b$  is the distance of the centre of the dislocation from the symmetrical configuration shown in Fig. 7.15. In this configuration ( $\alpha = 0$ ), the additional half-plane of atoms above  $A$  is midway between two half-planes below  $B$ ; in the simple cubic lattice there is a second symmetrical configuration ( $\alpha = \frac{1}{2}$ ) in which one half-plane on the expanded side, below  $B$ , is midway between two half-planes on the compressed side, above  $A$ .

Equations (31.1) and (31.4) enable the shear stress in the slip plane to be written as

$$X_{12}(x_1, 0) = B_e z_1 / (z_1^2 + \zeta^2), \tag{31.6}$$

where  $z_1 = x_1 + \alpha b$ . It also follows that the density of Burgers vector along the  $x_1$  axis in the continuous distribution model is  $b\zeta/\pi(z_1^2 + \zeta^2)$ , and this provides one method of finding the stress field  $X_{ij}$  throughout the crystal by integration of the field of an individual elastic dislocation over this distribution. This gives for the shear stress component, for example,

$$X_{12}(x_1, x_2) = \frac{B_e z_1 (z_1^2 + z_2^2 - 2x_2 z_2)}{(z_1^2 + z_2^2)^2}, \tag{31.7}$$

where  $z_2 = x_2 + \zeta$  for  $x_2 > 0$  and  $z_2 - \zeta$  for  $x_2 < 0$ . Comparing with eqn. (30.7) we see that this is almost identical with the field of a single elastic dislocation with its centre at  $x_1 = -\alpha b$ ,  $x_2 = \mp \zeta$  for the upper and lower half-crystals respectively.

The energy of the dislocation in this model consists of the elastic energy stored in the two parts of the crystal plus the misfit energy across the plane  $x_2 = 0$ , which is the effective core energy. The elastic energy per unit length is given by the integral  $\int X_{12}(x_1, 0) u_1(x_1) dx_1$  and within a cylinder of large radius  $r_e$ , this becomes

$${}^L W_s = \frac{1}{2} B_e b \ln(r_e/2\zeta). \tag{31.8}$$

For a small ribbon of unit length and width  $dx_1$ , the misfit energy is

$\left[ \int_0^{\Phi_1} -X_{12}(x_1, 0) d\Phi_1 \right] dx_1 = (\mu b^2/4\pi a) \{1 + \cos(4\pi w_1/b)\} dx_1$ , and on integrating this between limits  $x_1 = \pm \infty$ , the total misfit energy is found to be

$${}^L W_c = \frac{1}{2} B_e b. \tag{31.9}$$

Nabarro's expression for the self-energy of an edge dislocation is thus

$${}^L W_d = \frac{1}{2} B_e b \{ \ln(r_e/2\zeta) + 1 \}. \tag{31.10}$$

Comparing this with eqn. (30.15) we see that the energy is identical with that of an elastic dislocation in an isotropic continuum if the integration is cut off at a distance  $r_i = 2\zeta/\exp(1)$  from the centre of the dislocation. Alternatively, we may regard eqn. (31.8) as giving the elastic energy with the integration cut off at a core radius  $r_i = 2\zeta$ , and to  ${}^L W_s$  is added the core energy of eqn. (31.9). It is perhaps more logical to make the external surface of the cylinder of radius  $r_e$  stress free, as before, and when the work done by the additional

external forces on the cylinder is included, an energy  $B_e b/4$  has to be subtracted from the right-hand sides of eqns. (31.8) and (31.10). Comparing now with eqn. (30.16), we find the equivalent cut-off radius is  $r_i = 2\zeta/[\exp(1)]^{3/2}$ .

The same procedure may be used for a screw dislocation, although it does not then seem immediately obvious that it is correct to choose a particular plane on which to calculate the misfit energy. Eshelby (1949) showed, however, that if this assumption is made, identical results are obtained except that  $\zeta = \frac{1}{2}a$  and  $B_e$  is replaced by  $B_s$ . The results are independent of the choice of misfit plane.

The width of the dislocation given by eqn. (31.5) is very small and this implies a contradiction; the narrow core is a consequence of the sinusoidal approximation, but the approximation is valid only if the misfit is spread over many atoms. A more realistic law of force would correspond to the same initial slope but a smaller maximum force, and Foreman *et al.* (1951) developed a parametric solution for a family of force laws of more general type than (31.2). If the maximum force is reduced, so also is the energy of the regions of severe misfit in comparison with other regions, and the core therefore becomes wider. The form of the solution is similar to eqn. (31.4) but allows the dislocation width to be arbitrary; the product of the width and the maximum force is almost constant.

On p. 238 we referred to the constraining force tending to keep a dislocation in an equilibrium position in the lattice. The barrier to dislocation motion arises because of the periodic variation in the core energy of the dislocation as it is displaced, and the stress needed to move the dislocation over this barrier is generally called the Peierls stress. In the model we have just described, the effects of the discrete lattice are smoothed by the integration leading to eqn. (31.9), and in order to find the energy as a function of displacement (i.e. of the parameter  $\alpha$ ) it is necessary to sum the misfit energy over the atom rows in the cut. In the method used by Nabarro, this sum is evaluated separately over the top and bottom rows, and gives for the core energy

$${}^L W_c(\alpha) = \frac{1}{2} B_e b \{1 + 2 \cos(4\pi\alpha) \exp(-4\pi\zeta/b)\}. \quad (31.11)$$

The difference between the maximum and minimum values of this expression is the Peierls energy

$${}^L W_P = 2B_e b \exp(-4\pi\zeta/b) \quad (31.12)$$

and from the maximum slope of this expression, the corresponding Peierls stress is

$$X_P = (2\pi/b^2) {}^L W_P. \quad (31.13)$$

Using typical values for the elastic constants, eqn. (31.13) gives a stress of  $\sim 2-4 \times 10^{-4} \mu$ , i.e.  $\sim 10^7 \text{ N m}^{-2}$  or  $1 \text{ kg mm}^{-2}$ . This is about  $10^3$  times smaller than the theoretical stress to shear a perfect lattice, but is much larger than the observed low-temperature yield stresses of well-annealed single crystals of soft metals. However, it follows from (31.12) that the Peierls energy and stress are both very sensitive to dislocation width and the above values will be reduced when  $a$  is greater than  $b$  (as will usually be the case for slip on a close-packed plane) or when the force law is such as to give a greater width than that predicted by eqn.

(31.5). Foreman *et al.* confirmed that the exponential dependences of  ${}^LW_p$  and  $X_p$  on  $\zeta$  remain in their more general treatment.

Criticisms of the above calculation, and attempts to improve it, have been made by several authors. Huntington (1955) noted that the two symmetrical configurations of the dislocations have equal energies according to the above calculation, so that the wavelength of the Peierls force is  $\frac{1}{2}b$  instead of  $b$ . He pointed out that when the atoms are almost in registry in  $A$  and  $B$ , the forces on the atoms are nevertheless defined in terms of the large displacements (approaching  $\pm b/4$ ) from the reference positions, and he developed an alternative treatment in which the forces depend on the actual positions of the atoms. This gives a similar exponential dependence on the width of the dislocation, but now has the correct periodicity  $b$ ; the predicted Peierls stress has higher numerical values of  $\sim 10^{-2}$ – $10^{-3}\mu$ . Kuhlmann-Wilsdorf (1960) pointed out that the Peierls energy given by eqn. (31.12) is only a very small percentage of the core energy, and she suggested that the first-order effect has been lost in the Peierls–Nabarro calculation because of the equal energy of the two symmetrical configurations. She attempted to estimate the Peierls stress directly by summing the tangential forces on the atoms without calculating the energy, and obtained a result in terms of the core radius  $r_i$

$$X_p = (b/8r_i)X_{\text{crit}}, \quad (31.14)$$

where  $X_{\text{crit}}$  is the critical stress to cause slip in a perfect crystal. The exponential dependence on core width does not appear in this equation, but the numerical values of the Peierls stress are comparable with those obtained by Huntington.

Equation (31.12) is unsatisfactory not only in giving equal energies to the  $\alpha = 0$  and  $\alpha = \frac{1}{2}$  configurations, but also because these are the *maximum* energy configurations, whilst the asymmetric configuration ( $\alpha = \frac{1}{4}$ ) gives the minimum energy. This occurs because of the independent summations over the two planes  $A$  and  $B$ , and it seems physically more realistic to sum the interaction energy between pairs of atom rows in the two surfaces. This has been done numerically by Vitek (see Christian and Vitek, 1970) and produces the result that the minimum and maximum energy configurations now occur at  $\alpha = 0$  and  $\alpha = \frac{1}{2}$  respectively. The stress again decreases rapidly with the width of the core.

In recent years there have been many attempts to overcome the limitations of the Peierls–Nabarro model by the use of more realistic interatomic forces and by considering the atomic structure of a block of material and not just of the slip plane. In early work a discrete elastic approach (linear force laws between the atoms) was often used, and solutions for edge dislocations in a simple cubic lattice (Babuska *et al.*, 1960) and for screw dislocations in simple cubic (Maradudin, 1958) and diamond cubic (Celli, 1961) structures were obtained. However, discrete elastic models can be expected to be a good approximation only for relatively small deviations of the atomic structure from the defect-free equilibrium configuration, and thus do not work well for dislocation cores. Non-linear force laws, of the types discussed in Chapter 5, are thus required, and many computer calculations of the core structure of particular dislocations have now been published. One of the earliest of these calculations was made for edge and screw dislocations in sodium chloride (Huntington *et al.*, 1955, 1959) using electrostatic attractive and Born–Mayer repulsive forces; in this work,

rows of ions were treated as solid rods, but were not restricted in motion parallel to their length. The width of the edge dislocation was found to be similar to that given by the Peierls–Nabarro model.

Morse potentials were used by Cotterill and Doyama (1966) and Doyama and Cotterill (1966) to compute the structure of edge and screw dislocations in f.c.c. metals. The potentials were fitted to the measured elastic constants of copper and truncated to 179 neighbours. The dislocation cores were found to be undissociated when relaxation was allowed only parallel to  $b$ , but split into partials on  $\{111\}$  planes (see Sections 32 and 33) when full relaxation was allowed. Chang and Graham (1966) used a Johnson potential (eqn. 16.12) truncated to second neighbour interactions to calculate the core structure of an edge dislocation in a b.c.c. structure, and Yamaguchi and Vitek (1973) have calculated the structures of several non-screw dislocations in this structure, using general potentials of the Johnson type. These results all show the cores to be rather narrow and planar, as assumed in the Peierls–Nabarro model. However, most attention has been paid recently to core calculations for screw dislocations in b.c.c. structures, which are of especial interest since for reasons connected with their symmetry they have a non-planar structure (see pp. 311). Various calculations for the screw dislocation (Suzuki, 1968; Chang, 1967; Bullough and Perrin, 1968; Gehlen, 1970; Vitek *et al.*, 1970; Basinski *et al.*, 1970) are not in complete agreement, but some of the discrepancies arise not so much from the results themselves as from the difficulties of interpretation. The calculations have been extensively reviewed by Vitek (1974), but since they require a discussion of the geometry of the lattice, we defer further description to Section 33.

In an atomistic calculation it is not possible to follow the Peierls–Nabarro procedure of displacing the centre of the dislocation from its equilibrium configuration and calculating the corresponding energy. This is because each configuration is obtained by allowing the atomic positions to relax, so that only the energies corresponding to either stable or metastable configurations may be determined. The Peierls energy can thus not be calculated, since the configuration at which the energy is a (saddle-point) maximum is necessarily unstable. However, a correct procedure for the calculation of the Peierls stress is to apply external forces to the crystal block in small increments and to calculate the equilibrium configuration at each stage until a stress is reached at which the dislocation runs away. Such calculations were first made successfully for sodium under the action of shear stresses (Basinski *et al.*, 1971) and uniaxial stresses (Basinski *et al.*, 1972), using the potential described previously (eqn. 16.10). Under a pure shear stress the Peierls stress for a screw dislocation was found to depend on the orientation of the stress axis but to have a high minimum value of  $\sim 0.01\mu$ , whereas the corresponding stress in the h.c.p. lattice with the same interatomic potential was at least 25 times smaller. When uniaxial stresses were applied, the minimum resolved shear stress to move the b.c.c. screw dislocation was reduced, but only to  $0.007\mu$ . The results provide direct evidence that the crystal structure itself may be an important factor in fixing the magnitude of the Peierls stress. More recently, the distortion of the core and the motion of the screw dislocation in a b.c.c. structure under an applied shear stress has been calculated for a wide variety of effective interionic potentials of the type described by eqn. (16.12) (Duesbery *et al.*, 1973). The behaviour is very complex and is believed to account

for some of the anomalous results obtained from studies of the deformation of b.c.c. metals at low temperatures (Christian, 1970).

In b.c.c. metals and in many semi-conductors and non-metals, there is good experimental evidence for a high Peierls force, at least at low temperatures, but in close-packed structures the stress to move dislocations may be as low as  $10^{-5}$ – $10^{-6}$   $\mu$ . If the theoretical calculations are accepted, it is thus necessary to consider what additional effects may allow deformation to occur at such low stresses. One possibility considered by Dietze (1952) is that the Peierls stress is rather strongly temperature-dependent because of the effects of thermal vibrations. Unfortunately there is no evidence of a steep rise in stress at very low temperatures for the soft materials, and in the materials which have a strong temperature dependence of stress there is good evidence that this arises from thermally activated processes, i.e. from fluctuations rather than from an intrinsic temperature dependence of the barrier.

The effect considered by Dietze takes account of the fact that the atoms are not fixed points, but the dislocation line itself is essentially considered to be fixed in position. Kuhlmann-Wilsdorf (1960) pointed out that because of the atomic oscillations the exact position of the dislocation axis cannot be defined since certain modes of oscillation correspond to a displacement of the axis, and these modes can scarcely be correlated over long distances along the dislocation line. Her calculation suggested that this uncertainty in the position of the dislocation is much more important than the diffuseness in the positions of individual atoms, and that it may reduce the Peierls force to a very low value in f.c.c. metals, even at low temperatures. In materials such as diamond, this effect will not be operative until high temperatures are reached, whilst b.c.c. structures represent an intermediate case. In principle, direct temperature effects of this type (as distinct from problems involving thermal activation) may be calculated with atomistic models by adding kinetic energy to the atoms (see Gehlen, 1972), but the difficulties are considerable. A criticism of the Kuhlmann-Wilsdorf calculation is that over short lengths of dislocation line it requires coordinated motions of the atoms in opposite senses above and below the slip plane, and the amplitude ascribed to this mode is probably much too large.

Calculations of the Peierls force for straight edge and screw dislocations may be rather misleading, since the restraining force is much smaller when the dislocation does not lie along a close-packed direction of the lattice, and is zero for a completely arbitrary (irrational) direction. Shockley (1947) pointed out that we should thus expect restraining forces from the lattice only for straight dislocations of simple type, and he suggested that the force is probably zero even for these dislocations since the atomic configuration along a macroscopically straight line need not be constant. At sufficiently high temperatures, such a dislocation line will contain kinks at which it moves from one close-packed row in the slip plane to a neighbouring row. These kinks are analogous to the kinks in the steps on a crystal surface described in Section 18, and the dislocation line can move by sideways glide of the kinks along the line at a very much lower stress than is required to move the line forward as a whole. For a line with a macroscopic direction deviating slightly from a close-packed direction, there will be a certain concentration of geometrically necessary kinks; under the action of an applied shear stress, these will glide to the end of the line,

which will thus be turned into the close-packed direction. The continued motion of the line at finite temperatures then depends on whether thermal kinks form sufficiently rapidly, and the situation is again analogous to the growth of a crystal surface by addition of atoms at a step. Shockley suggested that thermal activation will provide kinks even at quite low temperatures, so that the resistance of the lattice to dislocation motion will be negligible except near 0 K. However, detailed consideration shows that the energy to produce kinks can be quite large when the Peierls barrier is large, and whilst Shockley's suggestion is probably true for close-packed metals at most temperatures, present evidence is that there is appreciable lattice resistance at quite high temperatures for b.c.c. metals and materials such as silicon and germanium.

Thermal energy may assist a dislocation line which lies in an equilibrium position to move forward to the next equilibrium position by providing an energy fluctuation which enables the applied stress to push a small portion of the line over the barrier. The critical condition thus involves the production of a double kink in an initially straight line, and processes of this kind were first discussed by Seeger *et al.* (1957) in connection with a low-temperature internal friction peak (Bordoni peak) in close-packed structures. For b.c.c. metals, measurements of the temperature and strain rate sensitivity of the flow stress, first made by Basinski and Christian (1960) and Conrad and Schoeck (1960), showed that the gliding dislocations are overcoming some obstacle to motion by means of thermal activation, and in these and many subsequent papers it has been suggested that the barrier is the lattice itself. The application of the rate theory of Chapter 3 to dislocation processes is very complex since stress is an independent variable, and a detailed description cannot be given here. The theory has been developed by many authors, and is reviewed, for example, by Hirth and Nix (1969), Evans and Rawlings (1969), Christian and Vitek (1970) and Kochs *et al.* (1975). It is now generally accepted that dislocation motion in b.c.c. metals is thermally activated, but some controversy has existed for many years about whether the resistance is due to the lattice periodicity or to dispersed interstitial impurities which interact strongly with dislocations (see Section 30). Stein and Low (1966) and others have provided experimental evidence in support of the impurity theory, but in a critical survey Christian (1968) concluded that it is very difficult to explain the low-temperature strength in this way. Whilst it is now clear that there are important impurity effects, present evidence indicates that the sessile equilibrium configuration of the screw dislocation (which is a form of Peierls–Nabarro force) is responsible for many of the low-temperature properties of b.c.c. metals. This conclusion is consistent with many electron-microscope investigations which show the low-temperature dislocation structure to consist of long, apparently immobile, screw dislocations.

In both the elastic solution and the Peierls–Nabarro approximation, the energy of an isolated dislocation line diverges, but the energy of a closed loop is finite. This suggests the possibility that the energy of formation of a loop within a perfect lattice might be supplied by thermal fluctuations assisted by an external stress. In view of our previous conclusion that the free energy of a dislocation is too high for it to be in thermal equilibrium with the lattice, however, it seems unlikely that this process is feasible. A detailed examination of the possibility of nucleating an opposite dislocation pair was made by Nabarro

(1947), using the Peierls–Nabarro expression for the self-energy of an edge dislocation pair, which differs from the elastic eqn. (30.29) only in the assumption that the normal stress  $X_{22}$  is zero. It seems preferable to treat this problem directly in terms of the nucleation of a closed loop of dislocation, using the approximate expression (30.30) (Cottrell, 1953). The increase in energy when a small loop is formed in the interior of a perfect crystal will be given by the dislocation energy (30.30) less the work done by the external shear stress  $X$  in forming the loop, i.e.

$$W_d = \frac{1}{2}\mu b^2 r \ln(r/r_i) - \pi X b r^2. \quad (31.15)$$

As the radius of the loop increases, the energy first increases and then decreases again. There is thus a critical size of loop given by

$$r_c = (\mu b / 4\pi X) \{\ln(r_c/r_i) + 1\} \quad (31.16)$$

at which the energy is a maximum

$$W_c = (\mu b^2 r_c / 4) \{\ln(r_c/r_i) - 1\}. \quad (31.17)$$

In terms of the theory of Chapter 3,  $W_c$  is now the activation energy for the formation of the loop, and the chance of a favourable thermal fluctuation is proportional to  $\exp(-W_c/kT)$ . Using typical figures, we find that with a critical radius of about 40 atom diameters, the applied stress is  $10^{-2}\mu$  and the activation energy about 20 eV. Thus even with this very high stress, the probability of spontaneously nucleating a dislocation loop is virtually zero. For still smaller loops, the activation energy decreases but the required external stress increases, and the general conclusion is that spontaneous formation of a slip dislocation in a perfect lattice (without stress raisers) will not take place unless the applied stress approaches the theoretical shear strength of the lattice. This conclusion does not necessarily apply to the formation of small loops of imperfect dislocation, e.g. at a twin interface, since the dependence of the activation energy on  $b^2$  makes thermal activation much more effective when  $b$  is small.

We now turn to consider the motion of a dislocation line in a crystal. In many circumstances the average velocity of a dislocation is limited by obstacles to its motion such as intersecting (“forest”) dislocations, impurity atoms, point defects, or the frictional drag of the lattice itself. When these obstacles are overcome with the aid of thermal fluctuations, the effective velocity is determined by the mean time of stay at each obstacle and not by the speed of the dislocation moving through the perfect crystal between obstacles. In a region of perfect crystal, however, we may expect the velocity of a dislocation to rise rapidly once it begins to move under an external stress, provided the lattice resistance to motion is small.

In the absence of energy dissipation, the work done by the stress field in moving the dislocation is equal to its gain in kinetic energy (we can use the concept of kinetic energy of the dislocation, which is a configuration of atoms, in much the same way as we use the concept of the force on a dislocation). It is obvious physically that the speed of sound sets an upper limit to the velocity of the dislocation, since motion of the latter corresponds to the propagation of an elastic discontinuity. We may thus anticipate that the energy

of a moving dislocation will tend to infinity as its velocity approaches that of the speed of sound. Actually, moving dislocations lose energy to the surrounding lattice, and this dissipation effectively produces a viscous drag, so that a limiting velocity is obtained under given stress conditions.

We begin by neglecting energy dissipation and consider a simple screw dislocation lying along the  $x_3$  axis, so that the strain field is given by eqn. (30.11). We wish to determine the energy of the dislocation when it moves with velocity  $u$  in the  $x_1$  direction. The simplest assumption is that the stress field of the dislocation is unaltered, so that if we use a set of coordinate axes which move with the dislocation line, the displacements are again given by eqn. (30.11). Referred to the usual stationary axes, we have

$$w_3 = (b/2\pi) \tan^{-1}\{x_2/(x_1 - ut)\}. \quad (31.18)$$

If we now calculate the energy of the moving dislocation, we find that it increases continually with  $u$  and is equal to twice its rest energy when  $u = c$ , the velocity of sound in the crystal. The solution is thus physically inadmissible since it permits dislocations to move with velocities greater than  $c$ . The error lies in the assumption that the strain field of a moving dislocation is identical with that of a stationary dislocation. As the motion proceeds, the material particles of the crystal are accelerated and then decelerated, and the inertial forces thereby introduced have to be included in the elastic equations of equilibrium. The situation is very closely analogous to the problem of a moving particle in the special theory of relativity; in deriving eqn. (31.18), we have applied a classical (Galilean) transformation instead of a Lorentz transformation.

Consider a small unit volume of material in the field of the moving screw dislocation. If the density is  $\rho$ , the inertial forces are equal to  $\rho(\partial^2 w_3/\partial t^2)$ , and the equations of elastic motion (11.7) become

$$\frac{\partial^2 w_3}{\partial x_1^2} + \frac{\partial^2 w_3}{\partial x_2^2} - \frac{1}{c^2} \frac{\partial^2 w_3}{\partial t^2},$$

where  $c = (\mu/\rho)^{1/2}$  is the velocity of transverse sound waves. We now introduce the new variables  $y_i$  where

$$y_1 = (x_1 - ut)/[1 - (u^2/c^2)]^{1/2}, \quad y_2 = x_2, \quad y_3 = x_3. \quad (31.19)$$

The equations of equilibrium then have solution

$$w_3 = (b/2\pi) \tan^{-1}(y_2/y_1). \quad (31.20)$$

Comparing with eqn. (31.18), we see that the only difference lies in the term  $[1 - (u^2/c^2)]^{1/2}$ . This corresponds to a reduction in the width of the strain field in the  $x_1$  direction, and may be compared to the Lorentz contraction of a moving measuring rod.

We have anticipated that the energy of the dislocation will tend to infinity as  $u \rightarrow c$ . Consider first the elastic energy. For a stationary dislocation, the elastic energy density in a small volume  $dx_1 dx_2 dx_3$  is given by

$$dW_s = \frac{1}{2} \mu \{(\partial w_3/\partial x_1)^2 + (\partial w_3/\partial x_2)^2\} \quad (31.21)$$

and the total elastic energy is

$${}^LW_s = (\mu b^2/4\pi) \ln(r_e/r_i) \quad (31.22)$$

per unit length as in eqn. (30.18). For the moving dislocation, (31.21) still gives the elastic energy density, and in terms of the new variables  $y_i$  we have

$${}^LW'_s = \iiint \frac{1}{2} \mu \{ (\partial w_3 / \partial y_1)^2 [1 - (u^2/c^2)]^{-1/2} + (\partial w_3 / \partial y_2)^2 [1 - (u^2/c^2)]^{1/2} \} \cdot dy_1 dy_2 dy_3.$$

This integral may be evaluated since the displacements are now the same function of  $y_i$  as they were of  $x_i$  for a stationary dislocation. By symmetry, each term in (31.21) contributes equally to the integral in (31.22); it follows that we may write

$${}^LW_s = \frac{1}{2} {}^LW_s [1 - (u^2/c^2)]^{-1/2} + \frac{1}{2} {}^LW_s [1 - (u^2/c^2)]^{1/2} = {}^LW_s [1 - (\frac{1}{2}u^2/c^2)] [1 - (u^2/c^2)]^{-1/2}. \quad (31.23)$$

In addition to the elastic energy, the moving dislocation now possesses kinetic energy. The energy of a small volume  $dv$  is  $\frac{1}{2} \rho (\partial w_3 / \partial t)^2 dv$ , so that we have

$${}^LW'_k = \iiint \frac{1}{2} \rho (\partial w_3 / \partial y_1)^2 u^2 [1 - (u^2/c^2)]^{-1/2} \cdot dy_1 dy_2 dy_3 = {}^LW_s (u^2/2c^2) [1 - (u^2/c^2)]^{-1/2} \quad (31.24)$$

substituting  $\rho/\mu = 1/c^2$ . Adding together eqns. (31.23) and (31.24), the total energy of the dislocation per unit length is found to be

$${}^LW'_d = {}^LW_s [1 - (u^2/c^2)]^{-1/2}, \quad (31.25)$$

where  ${}^LW_s$  is the elastic energy of the stationary dislocation. This result is exactly parallel to the formula for the energy of a moving particle in the theory of relativity. As  $u \rightarrow c$ , the energy approaches infinity.

The simple analogy between the formulae of the special theory of relativity and those applicable to a moving screw dislocation is a little misleading. For edge dislocations, the situation is more complicated since the displacements are not all perpendicular to the direction of motion, and the velocities of transverse and longitudinal sound waves have both to be considered. Although the equations are more complex, however, the general results remain valid. Eshelby (1949) showed that  $c$  is also the limiting velocity for a uniformly moving edge dislocation, but the energy diverges as  $\{1 - (u^2/c^2)\}^{-3/2}$ . An interesting result is that the shear stress in the glide plane changes sign for  $c_R < u < c$ , where  $c_R \approx 0.9c$  is the velocity of Rayleigh waves. Thus a very fast edge dislocation attracts other edge dislocations of like sign and repels dislocations of opposite sign.

For velocities  $u \ll c$ , an expansion of eqn. (31.25) enables the energy to be written

$${}^LW'_d = {}^LW_s + \frac{1}{2} m_0 u^2, \quad (31.26)$$

where

$$m_0 = {}^LW_s/c^2 = (\rho b^2/4\pi) \ln(r_e/r_i) \quad (31.27)$$

may be regarded as the "rest mass" of unit length of a screw dislocation. The corresponding expression for the rest mass of an edge dislocation includes a factor  $\{1 + (c/c_1)^4\}$ , where  $c_1$  is the longitudinal sound velocity.

From eqn. (31.25) we see that dislocations moving with velocities which are an appreciable fraction of the velocity of sound possess energies which are much larger than the self-energy of a stationary dislocation. When  $u \approx 0.8c$ , for example, the rest energy is doubled, but for  $u \approx 0.1c$ , the difference is very small. A distinction has sometimes been made between dislocations which possess additional energy of motion comparable to their rest energies (fast dislocations), and those which do not (slow dislocations). The division is fairly sharp because of the form of equation (31.25). In principle, fast dislocations should be able to use their extra energy to overcome obstacles, and possibly to create new dislocations, and this was the basis of an early theory of dynamic multiplication of dislocations during mechanical deformation (Frank, 1951). The available experimental evidence at present is that dislocations do not attain velocities of the order of half the velocity of sound, either because of the presence of obstacles of various kinds, or because of the energy dissipation from a moving dislocation. However, the effective stress on an interface dislocation (see pp. 291, 363) during a phase transformation can be very high, and it is not yet certain that fast dislocations do not play a role in transformation phenomena.

Although the speed of sound appears as a limiting velocity in the above equations, Eshelby (1956b) has investigated the formal possibility of supersonic dislocations. Instead of (31.19), the transformation

$$y_1 = (x_1 - ut)/[(u^2/c^2) - 1]^{1/2} \quad (31.28)$$

is used and leads to the hyperbolic form

$$\frac{\partial^2 w_3}{\partial y_1^2} - \frac{\partial^2 w_3}{\partial y_2^2} = 0 \quad (31.29)$$

with the general solution

$$w_3 = \Omega_a(y_1 + y_2) + \Omega_b(y_1 - y_2). \quad (31.30)$$

Here  $\Omega_a, \Omega_b$  are plane waves of velocity  $c$  inclined to the  $x_1$  axis at angles of  $\tan^{-1} \pm [(u^2/c^2) - 1]^{1/2}$ . They represent outward radiation of energy from the moving disturbance and are similar to supersonic shock waves.

Eshelby showed that steady supersonic motion is only possible if energy is continuously supplied by atomic readjustment in the glide plane. In the Peierls-Nabarro model, for example, the solution implies that  $\alpha = 0$  before the disturbance has passed and  $\alpha = \frac{1}{4}$  after it has passed; i.e. the misfit energy across the glide plane is a maximum before passage of the supersonic dislocation and a minimum after it. In detailed studies, Weertman (1967) has confirmed that a dislocation moving on a glide plane which cannot contribute energy by atomic readjustments is unable to move supersonically. Clearly supersonic motion can not arise in ordinary glide, but it may occur for the special kind of dislocations which are essentially steps in twin boundary or martensitic interfaces (see Section 32). It may be significant that audible clicks are often produced by the formation of twins or martensite plates.

We now briefly consider some of the ways in which energy may be lost from a moving dislocation. The least important, at least until quite high velocities are attained, is thermoelastic damping. As a dislocation moves, the hydrostatic stresses at any point in the material are altered, and these changes produce corresponding changes in the local temperature. Heat flow takes place as a result of the small temperature gradients thus created, and this is one source of energy loss. Estimates show that the effect is small at most velocities. A second source of loss is the interaction of a dislocation with the natural vibrations of the lattice. In effect, the dislocation scatters sound waves which impinge upon it, so that these waves offer a resistance to the motion of the dislocation. Leibfried (1950) considered this problem and concluded that the limiting velocity of a dislocation would be only  $0.07c$  for an applied stress of  $5 \times 10^6 \text{ Nm}^{-2}$ . Later work by Nabarro (1951) showed that Leibfried had incorrectly calculated the scattering cross-section, the contribution of long elastic waves to the resistance being overestimated. A good estimate of this effect is very difficult to obtain, but Nabarro's tentative conclusion was that fast dislocations might be produced at very high stresses ( $\sim 10^{-3} \mu$ ), but it is improbable that there will be any fast dislocations at the much lower stresses required to deform annealed single crystals.

A third source of energy dissipation arises from the periodic nature of the strain field of the dislocation. The variation in the form of the dislocation as it moves from one equilibrium position in the lattice to the next has already been considered; this variation leads to the radiation of elastic waves out from the dislocation. Nabarro also estimated the magnitude of this radiation damping; it is smaller than the scattering of sound waves, but not negligible. His calculations suggest that an applied stress ten times greater than the yield stress of soft single crystals is probably required to overcome this resistance sufficiently to produce fast dislocations.

The dynamic properties of dislocations form one of the most difficult parts of the formal theory, and many other mechanisms of energy dissipation have also been suggested (see Nabarro, 1967, and Hirth and Lothe, 1968, for further description). The magnitudes of these effects are very difficult to estimate, and as Eshelby has remarked, it is not always clear that they are separate from one another. Some experimental observations which were formerly thought to require fast dislocations have been explained in other ways, and it remains uncertain whether or not fast dislocations are encountered in normal deformation conditions.

We conclude this section by briefly discussing experimental measurements of individual dislocation velocities which were first made by Johnston and Gilman (1959) and Stein and Low (1960) who used an etch-pit technique to reveal the positions of a dislocation before and after a stress pulse of known duration. The results are usually fitted to empirical equations of form

$$u = u_0(\tau/\tau_0)^m \quad (31.31)$$

or

$$u = u_0 \exp(-\tau_0/\tau), \quad (31.32)$$

where  $\tau$  is shear stress; the parameters in these equations are probably functions of temperature and possibly also of stress, i.e. they are not representative of the correct functional relations. The first equation is usually applied only over the range of velocities which

correspond to obstacle-limited motion, and the true functional relation between velocity and stress is then probably better represented by the rate equation

$$u = \nu d \exp(-\Delta_a G/kT), \quad (31.33)$$

where  $\nu$  is the attempt frequency of the dislocation held up at an obstacle,  $d$  is the distance moved found after a successful activation and  $\Delta_a G$  is the free energy of activation (cf. Chapter 3). These velocity measurements are usually consistent with those deduced from macroscopic measurements of the stress and temperature sensitivity of the strain rate, which is given by

$$\dot{\epsilon} = \rho b u, \quad (31.34)$$

where  $\rho$  is the total length of mobile dislocation in unit volume (see p. 313). We have already mentioned the rate theory analysis of flow stress data (p. 280), but we should emphasize that the applicability of rate theory to either individual velocities or to macroscopic strain rates is not universally accepted.

Equation (31.32) was first used by Gilman (1960) and was developed on the basis of a rather specific and oversimplified rate theory model. As an empirical equation, however, it has been claimed to represent the relation between velocity and stress over most of the range which can be measured (in the case of LiF from  $10^{-8}$  to  $10^{+25}$  m s<sup>-1</sup>). At the highest velocities, which correspond to one-tenth of the velocity of sound, the motion is probably limited by energy dissipation rather than by the time of stay at obstacles, and such velocities correspond to very high strain rates in macroscopic experiments.

### 32. PARTIAL DISLOCATIONS, TWINNING DISLOCATIONS, AND TRANSFORMATION DISLOCATIONS

Dislocations which are "perfect" are surrounded entirely by good crystal. We now consider whether any physical significance can be assigned to dislocations with Burgers vectors not equal to repeat vectors of the lattice. In principle, such dislocations can be made by the virtual process described on p. 257; we cut a perfect crystal along some surface having the dislocation line as its limit, move the two cut faces by a non-lattice vector (adding or removing thin layers of material as necessary), and then rejoin. An important difference is at once apparent. After displacement, the lattices of the two cut portions will no longer match, and when rejoined, the cut surface will generally be a region of high energy. An imperfect, or partial, dislocation is thus associated with a surface of misfit, and the possible virtual processes leading to the same imperfect dislocation are not equivalent, since the energy of the misfit surface will depend on its orientation relative to the lattice. In most, and possibly all, orientations, this energy will be very large, and the existence of such irregularities within a crystal is most improbable.

Imperfect dislocations are thus to be expected only when misfit surfaces of low energy can exist. Such surfaces are provided by the stacking faults or translational twin interfaces described on pp. 122-7. Unlike ordinary twin boundaries, a translational twin interface may end within the crystal, and the edge of the fault is then an imperfect dislocation. If there

is no marked discontinuity in the elastic properties at the stacking fault, there will be no strong asymmetry in the stress field of the imperfect dislocation, and the previous elastic theory may be used. Since the stacking fault is planar, the imperfect dislocation is properly regarded as a two-dimensional fault.

The Burgers vector of an imperfect dislocation may be defined formally in exactly the same way as for a perfect dislocation (p. 244) with one or two additional conditions (Frank, 1951). The stacking fault is strictly a region of bad crystal, and to avoid ambiguity the Burgers circuit must begin and end on the fault and not at an arbitrary point. This gives an unambiguous path in the reference lattice for intrinsic faults (p. 123), since the path in the real crystal is entirely in good crystal. For extrinsic faults, we must also specify that the interpolated layer is crossed normally, and in the reference lattice we take an equal distance normal to the corresponding plane.

Suppose we have a strip or a loop of stacking fault, bounded by imperfect dislocations on either side. The sum of the Burgers vectors of opposite bounding dislocations must clearly be a lattice vector in order to leave the structure undisturbed outside the fault. If this sum be zero, the partial dislocations are opposites and may annihilate each other by moving together. This is analogous to the formation of a loop of perfect dislocation by local slip, and we may assign a single Burgers vector to the whole dislocation by specifying that the loop is to be traversed in a certain sense (see p. 245). When the two edges of a stacking fault are not opposites, their Burgers vectors add to give a non-zero lattice vector. If the two vectors have a component in the same direction, the lines will repel each other and the edges of the stacking fault move apart. As they do so, the elastic energy of the partial dislocation decreases, but the area of the fault increases, so that there is an equilibrium separation at which the total energy (free energy) is minimized. This separation may be estimated by balancing the elastic repulsive force between the partials and the attractive force which results from the surface tension of the stacking fault. In metals and alloys of low stacking fault energy, ordinary lattice dislocations are dissociated into ribbons of stacking fault in this way whenever they lie in low fault energy planes. The combination of two partial dislocations and a connecting ribbon of stacking fault is described as an extended dislocation.

The two-dimensional nature of the fault attached to an imperfect dislocation imposes additional restrictions on the motion of the dislocation. Suppose the Burgers vector lies in the composition plane of the stacking fault. Gliding motions of the dislocation in this plane are then possible; as they take place, the area of stacking fault increases or decreases, and the corresponding change in free energy imposes an additional force on the dislocation. Gliding motions on any other plane are not possible, even if the dislocation assumes screw orientation, since the surface of misfit would then be extended out of the composition plane and so would have high energy. An exception to this rule occurs if the Burgers vector of an imperfect dislocation is common to two fault-planes. A dislocation lying in one plane can then assume screw character over part of its length, and thus extend into the second plane. The necessary condition is obtained in the b.c.c. structure, and may be important in connection with the formation of mechanical twins. In the close-packed lattices, however, the possible imperfect dislocations bounding stacking faults each lie in only one of the composi-

tion planes, and their gliding motions are thus confined to these planes.<sup>†</sup> Climbing motions of the dislocation line outside its composition plane would also lead to the production of a high energy interface, and so are also impossible.

Now consider the dislocation line to have a Burgers vector which does not lie in the composition plane of the low energy stacking fault which it produces. Gliding motions are then not possible in the composition plane, since the glide plane must contain the Burgers vector, and gliding motions out of the composition plane are forbidden as described above. Such imperfect dislocations are thus unable to glide at all; they are called "sessile" dislocations, in contrast to normal dislocations which are "glissile". Sessile dislocations may climb by atomic diffusion, provided they remain in the same composition plane. All imperfect dislocations are thus constrained to move only in the composition plane of the low energy fault; the motion requires atomic diffusion if the Burgers vector does not lie in this composition plane, and occurs by glide if it is in this plane.

Imperfect dislocations exist only at the edges of low energy stacking faults. It is possible to give a description of any internal misfit surface in dislocation terms, though in some cases this has formal rather than physical significance. We shall consider internal boundaries in the next chapter, but it is convenient here to describe the properties of a defect closely related to imperfect dislocations. We have seen in Chapter 2 that mechanical twinning may be described as a simple shear of the lattice along a composition plane  $K_1$ . Figure 7.16 shows diagrammatically a twin of type I in which the simple shear is produced by the lattice planes each slipping a certain distance over the plane beneath. On the plane  $AB$ , this slipping motion has extended over part of the plane only, so that the actual boundary between the twin and the parent crystal extends from  $A$  to  $P$  and then steps on to the next lattice plane.

It is clear that around the step at  $P$ , which runs perpendicular to the figure, there is a region of high strain energy. This step is the limit of a region of slip, and hence has all the properties normally associated with a dislocation. It is called a twinning dislocation, since it exists only in a twin boundary and not in the interior of a crystal. The Burgers vector of a twinning dislocation is not a lattice vector of the parent or twinned lattice, but is equal to the distance moved by each lattice plane in the twinning shear. Thus for a shear  $s$ , the Burgers vector of the twinning dislocation is

$$\mathbf{b} = ds, \quad (32.1)$$

where  $d$  is the step height. The minimum Burgers vector is fixed by the smallest repeat distance between lattice planes, and for a given  $K_1$  plane we expect all the twinning dislocations to have minimum Burgers vectors, since multiple steps will tend to dissociate into single steps. However, this conclusion is only valid if the atomic configuration at the interface is identical in the two adjacent lattice  $K_1$  planes connected by a step of minimum height. As mentioned on p. 57, the crystallography of twinning is sometimes such that the lattice deformation  $S$  does not relate all of the lattice vectors of the two structures, and this implies

<sup>†</sup> This does not mean that two stacking faults on different planes of a f.c.c. structure, for example, may not meet along the line of intersection of the two planes. The difference is that the line of intersection would itself be an imperfect dislocation line, whereas the line of intersection is not a dislocation if the imperfect dislocation can glide in both planes.

that the interface structure is repeated only at distances corresponding to an integral multiple of the spacing of  $K_1$  planes. A detailed discussion of this complication is given in Part II, Chapter 20; if the differing interface configurations have very different energies, the actual step heights may correspond to the distance at which the lowest energy configuration repeats. Such a step is called a zonal twinning dislocation (Kronberg, 1959, 1961; Westlake, 1961).

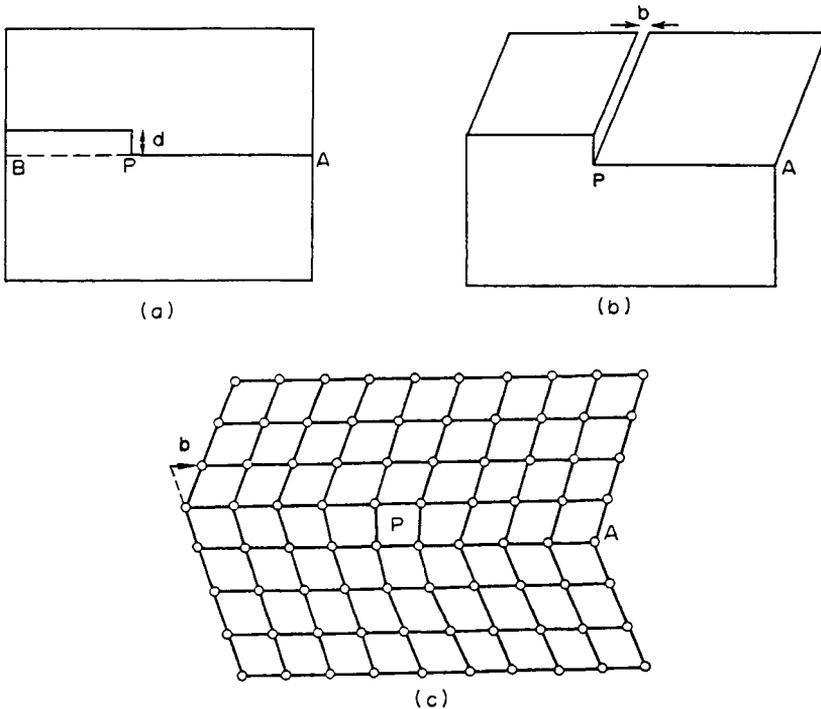


FIG. 7.16. To illustrate the formation of a twinning dislocation (after Bilby and Christian, 1956). When the upper half of the crystal shown in (a) is sheared into twin orientation, a fissure appears because of the step at  $P$ . When this is rewelded to produce a twin boundary with a step, the discontinuity at  $P$  has the characteristics of an edge dislocation. The approximate atomic configuration in a simple structure is shown in (c).

The Burgers vector of the twinning dislocation is always smaller than a lattice vector. The dislocation can glide in its composition plane (the  $K_1$  plane), and as it does so, the amount of one orientation grows at the expense of the other. Macroscopic growth of a twin could obviously result from a mechanism which allows a twinning dislocation to move through a whole series of parallel lattice planes. Gliding motions of the twinning dislocation outside the  $K_1$  plane, or climbing motions, produce a high energy interface (an incoherent twin boundary), and are thus impossible under normal stresses.

The step in the twin boundary of Fig. 7.16 is perpendicular to the  $\eta_1$  direction, and is thus an edge-type twinning dislocation. Obviously dislocations of screw and mixed types are also possible, and closed dislocation loops may be formed. Figure 7.17 illustrates both

edge- and screw-type twinning dislocations in a face-centred tetragonal structure with a  $\{101\}$  twinning plane. Mechanical twins often have a lenticular shape, and the progressive twinning may be described formally in terms of closed loops of twinning dislocation with decreasing radius as we move outwards from the centre plane of the lens.

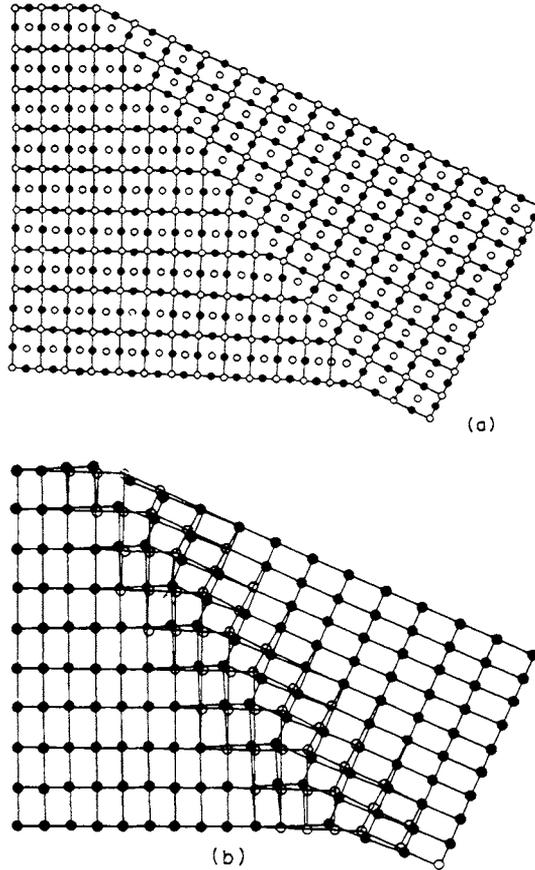


FIG. 7.17. Two edge-type twinning dislocations (a) and one screw-type twinning dislocation (b) in the  $\langle 101 \rangle$  twin boundary of a tetragonal structure (after Basinski and Christian, 1954).

The plane of the figures is  $\{010\}$ . Open and filled circles represent two successive atom layers projected on this plane. To avoid confusion, (b) is drawn for a simple tetragonal structure, but may readily be related to the face-centred tetragonal structure of (a).

The Burgers vector of a twinning dislocation may be defined formally by using a reference lattice which represents not a single crystal but a crystal and its coherent twin meeting on a  $K_1$  plane. The reference lattice is then multi-valued; it consists of two interpenetrating lattices since the composition plane can be any one of the infinite series of lattice planes parallel to  $K_1$ . If the Burgers circuit is made round a step in the real boundary, the corresponding circuit in the reference lattice will have a closure failure which will be a vector parallel to  $s$  and connecting a point of one of the interpenetrating lattices to a neigh-

bouring point of the other. This is the formal version of the definition of a twinning dislocation used above, but we note especially that it is valid whether or not the twin has formed by a physical deformation. A step in a twin boundary has the characteristics of a dislocation, even if the twin itself were formed by individual thermally activated atomic movements, and not by shear.

In Section 9 we generalized the concept of simple shear to that of invariant plane strain, which produces a new lattice. We may now repeat the above arguments by imagining two lattices which are related by an invariant plane strain, and which meet along some composition plane which we shall assume, for the present, to be rational. If the composition plane contains a step where it leaves one lattice plane and moves to a neighbouring plane, this discontinuity will have properties very similar to those of the twinning dislocation described above; it is described as a transformation dislocation (Bilby, 1953). As a transformation dislocation glides in its composition plane, atoms are transferred from one phase to another. We saw on p. 60 that the displacements in an invariant plane strain are equivalent to the combination of a simple shear on the composition plane and a uniaxial expansion or contraction normal to this plane. The Burgers vector of the transformation dislocation, defined formally by using a multivalued reference lattice comprising the interpenetrating lattices of the structures, will obviously be of this form, and will be equal to the vector  $\mathbf{e}$  of eqn. (9.10) multiplied by the height of the step. We may also envisage the situation when the invariant plane strain  $\mathbf{S}$  relates only some fractions of the lattice points in the two structures; this leads to the concept of multiple step heights or "zonal transformation dislocations".

According to the above descriptions, the transformation dislocation moves in a fixed composition plane which need not contain its Burgers vector. The usual restriction that glide motion of a dislocation is only possible in a plane containing its Burgers vector is relaxed because the dislocation lies in the boundary between two lattices which possess identical but differently spaced composition planes.

We next consider what happens when a dislocation lying in the parent lattice meets a coherent twin boundary. If this boundary were the free surface of the crystal, the dislocation would produce a step of height equal to the component of the Burgers vector normal to the boundary, and running from one end of the dislocation to the edge of the surface, or to the end of another dislocation. We may treat the twin by making an imaginary cut along the coherent composition plane and introducing the dislocation into the parent crystal, thus producing the step. Clearly the parent and twin can only be rejoined if there is also an incomplete step on the twin surface of the composition plane, and there must therefore be a dislocation line in the twin crystal. There is thus a node in the interface formed from (at least) one dislocation line in each crystal and a step, or twinning dislocation, lying in the interface. If the node contains three lines, as above, conservation of the Burgers vectors requires that the dislocations in the two lattices have equal components normal to the twinning plane (equal and opposite if the convention of eqn. (29.2) is used), and this component gives the height of the step. It is, in fact, geometrically obvious that the Burgers vector of the dislocation in the twin is produced from that of the dislocation in the parent by operation of the twinning shear; if we start with a single crystal containing a dislocation

and convert part of it into twin orientation by means of a simple shear, a node of the above type is formed naturally at the point where the dislocation crosses from the sheared to the unsheared portion. The only exception is for a dislocation with its Burgers vector parallel to the composition plane. The vector is then undistorted by the twinning law, there is no step in the composition plane, and the dislocation simply bends in passing from one lattice to the other.

Provided the twinning law is such that a simple shear  $S$  can be found to relate all the lattice sites of the two crystals, it follows that if the dislocation in one lattice has a Burgers vector which is a repeat vector of that lattice, so also does the dislocation of the other lattice. However, if  $S$  converts only some fraction of the lattice sites of one crystal to those of its twin, this condition may or may not be fulfilled. In the general case, both dislocations will have lattice Burgers vectors if the component of Burgers vector normal to the interface corresponds to the step height of a zonal twinning dislocation, rather than to that of an elementary twinning dislocation (Saxl, 1968); this is discussed further in Part II, Chapter 20. This result has an immediate application to the problem of glide of a dislocation across a coherent twin boundary. Unless the dislocation crosses in pure screw orientation, it must leave a step of height equal to the normal component of its Burgers vector along the line on which it crosses, with an interface node at each end of this line. If the step height is a repeat distance of the interface periodicity, the dislocation can continue to glide in a slip plane defined by its new Burgers vector and by the line of intersection, or step.

In other cases, the dislocation cannot glide into the new lattice without leaving a fault, but a group of dislocations in one lattice may form a single dislocation which glides into the other. It should be noted that anomalous slip planes and directions may be produced by this mechanism, as has been confirmed experimentally for slip across twin boundaries in zinc (Tomsett and Bevis, 1969). The new Burgers vector is related to the net Burgers vector of the group of dislocations by  $S$ , and the two slip-planes meet edge to edge along the line of the zonal twinning dislocation.

The above results may again be generalized to two lattices related by an invariant plane strain. In this case, the dislocations in the lattices will produce steps of unequal height on the composition plane, the difference being the expansion or contraction normal to the composition plane which is produced by the normal component of the Burgers vector of the transformation dislocation. A formal proof of the possibility of constructing this kind of node is as follows. If the relation between the lattices is represented by  $v = Su$ , it was shown on p. 60 that for an invariant plane strain, the tensor  $S$  must have matrix representation  $I + ev'$  where  $v$  is the normal to the invariant plane, and  $e$  is a constant vector which gives the direction of movement of the lattice points. It follows that we can construct a node of dislocation lines with Burgers vectors having representations  $u$ ,  $e v' u$ , and  $-v$ , since their vector sum is necessarily zero. Now if  $S$  relates all the lattices points and  $u$  represents the Burgers vector of a perfect dislocation in the first crystal,  $-v$  is a lattice vector of the second crystal and so represents the Burgers vector of a perfect dislocation in that crystal, whilst the vector  $e v' u$  represents a transformation dislocation. In the more general case, both dislocations in the lattices will have lattice Burgers vectors if the step represents a zonal transformation dislocation, and it is easily seen that all nodes between lattices are of the

above form. If two lattices are not related by an invariant plane strain, they do not possess a low-energy composition plane, and the concept of transformation dislocation is then not so useful. In some circumstances it may be possible to define a step in a semi-coherent interface as a transformation dislocation, but it is probably not possible to define unambiguously what happens to a dislocation line when it meets an incoherent boundary between two structures. The formation of a generating node, as discussed above, is illustrated in Fig. 7.18.

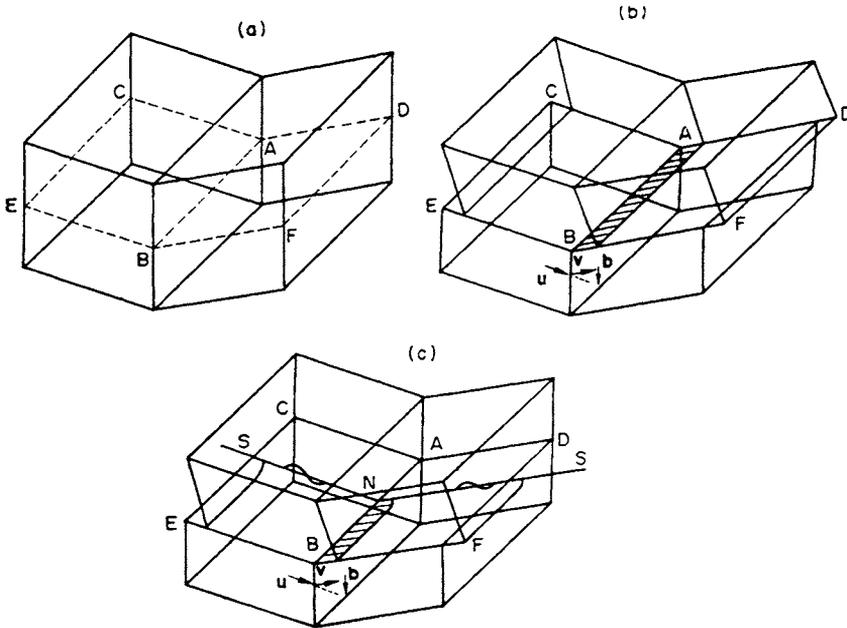


FIG. 7.18. To illustrate the formation of a generating node between two lattices (after Bilby and Christian, 1956). (a) Two coherent lattices (related by an invariant plane strain) with corresponding slip planes outlined. (b) Effect of slipping upper half of both crystals over lower half. The slip vector  $v$  of one crystal is produced from the corresponding vector  $u$  of the other crystal by the invariant plane strain. A transformation dislocation of Burgers vector  $b = v - u$  is formed along  $AB$ . (c) The slip now extends only from  $EBF$  to  $SNS$ , which is a screw dislocation in both crystals. The transformation dislocation extends from  $B$  to  $N$  where  $SS$  intersects the boundary. The boundary surface is now helicoidal, and rotation of  $NB$  about  $SNS$  generates one lattice from the other.

The concepts of twinning dislocation and transformation dislocation were introduced above only for rational composition planes, where the atomic arrangement in the neighbourhood of the composition plane is unambiguous and a clear meaning can be given to a step connecting two lattice planes. We should note, however, that a step produced by any dislocation, even the shortest perfect dislocation, crosses many lattice planes of type  $(h_1 h_2 h_3)$  if the indices  $h_i$  are not small. As the indices increase, the number of lattice planes crossed by the step increases and becomes infinity for the limiting case of irrational planes. The position of the interface is then no longer clearly defined, but the above arguments show that the concepts of twinning and transformation dislocations may still be useful.

If a dislocation in one lattice meets such an interface, e.g. in twinning of the second kind, it will produce a step equal to the component of the Burgers vector normal to the irrational plane, and this step is still a dislocation in the sense used above. The definition is such that if a given lattice point is in or near the composition plane, a lattice point related to it by the Burgers vector of the dislocation producing the step will lie in or near the new position of the composition plane (extended if necessary). A rather similar discussion has been given by Cahn (1960) of the existence of steps on the (probably diffuse) liquid–solid interface (see p. 166). In this further extension of the concept of a dislocation, we have now included discontinuities which can glide only in a given irrational plane. There are no further conditions on the dislocations forming the node in the boundary surface, and the dislocations in the two crystals may, in principle, be either perfect or imperfect. Nodes of this type have been called “generating nodes” by Bilby (1953); they are further discussed in Section 87.

The above description shows that despite the formal similarity, there are differences between the concepts of ordinary dislocations and those of dislocations which are present only in boundaries. As will become clear in the next chapter, a boundary may itself be regarded as a closely spaced array of perfect dislocations of the lattice, and a twinning dislocation is then a crossing dislocation or step in this dislocation array. Frank and van der Merwe (1949) proposed the name second-order dislocations for steps of this kind; another example of a second-order dislocation is the jog where a dislocation leaves one close-packed plane and moves to another (see p. 252). In the same way, a stacking fault may be regarded as a double dislocation array, and the associated imperfect dislocation is then the crossing dislocation which terminates this double array. The exact description is clearly a matter of convenience, and these rather fine details of the theory will not concern us further.

### 33. THE STABLE DISLOCATIONS OF THE COMMON METALLIC STRUCTURES

The Burgers vector of a perfect dislocation may be any of the vectors  $\mathbf{u}$  of eqn. (5.1), so that in principle there is an infinity of possible dislocations for any crystal structure. However, the strain energy increases rapidly with the length of the Burgers vector, and large dislocations will thus tend to dissociate spontaneously into lines with smaller Burgers vectors. By examining the change in energy resulting from a given dissociation, we can determine whether or not it will occur spontaneously, and in this way we arrive at a finite number of dislocations which are stable in each lattice.

According to the formulae of Sections 30 and 31, the elastic energy per unit length of dislocation is proportional to  $b^2$ , and to a small factor between 1 and  $(1-\nu)^{-1}$  which depends on the orientation of the line itself. In real crystals, the anisotropy of elastic properties and the variation of core structure may increase the orientation dependence, and in extreme cases this leads to the result that some orientations of a dislocation line cannot exist. We consider only effects due to elastic anisotropy, since the variation of core energy with orientation is generally unknown. The concept of line tension shows that in the absence of a stress field, a dislocation fixed at two points of an elastically isotropic crystal will take the form of a straight line joining these points. However, if the orientation dependence of

the line energy  ${}^LW_s$  is sufficiently large, the line tension as defined by eqn. (30.34) may become negative, and the straight dislocation line may then lower its energy by adopting some kind of zigzag configuration. This effect was first predicted for lithium by de Wit and Koehler (1959).

The equilibrium shape of a closed dislocation loop is given by a two-dimensional Wulff-plot type of construction (see p. 155), as first pointed out by Mullins (see Friedel, 1964). Head (1967) has used Frank's inverse Wulff plot (i.e. in the dislocation case a plot of  $\{{}^LW_s(\varphi)\}^{-1}$  vs.  $\varphi$ ) to examine the problem of dislocation line stability in detail. If there is any region in which part of this plot lies inside a common tangent to the plot, the corresponding orientations of the dislocation line are unstable and may change spontaneously to a V-shaped, Z-shaped, or more complex zigzag in which the component directions of the line correspond to the points of contact of the common tangent. This is a more general condition than that of negative line tension, and predicts wider ranges of instability.

Except for certain simple orientations, the energy  ${}^LW_s$  must be calculated numerically. Head has shown that in cubic crystals the instability ranges depend on the usual elastic anisotropy factor (Table IV, p. 76) and also on a second factor  $(c_{11} + 2c_{12})/c_{44}$ . The experimental values are such that ordinary lattice dislocations, but not necessarily partial dislocations, are stable in all orientations of f.c.c. metals, but there are forbidden dislocation orientations for b.c.c. metals. However, we must first determine what are the stable Burgers vectors in the common metallic structures, and for this purpose we assume the validity of the simple criterion that the energy varies as  $b^2$ .

In a dislocation reaction

$$\mathbf{b}_3 = \mathbf{b}_1 + \mathbf{b}_2$$

the dislocation with Burgers vector  $\mathbf{b}_3$  will be unstable if  $b_3^2 > b_1^2 + b_2^2$ . Since  $b_3^2 = b_1^2 + b_2^2 + \mathbf{b}_1 \cdot \mathbf{b}_2$ , the dislocation will dissociate spontaneously if  $\mathbf{b}_1 \cdot \mathbf{b}_2$  is positive (angle between  $\mathbf{b}_1$  and  $\mathbf{b}_2$  less than  $90^\circ$ ) and will be stable if  $\mathbf{b}_1 \cdot \mathbf{b}_2$  is negative (angle greater than  $90^\circ$ ). There is no first-order change in the energy if  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are perpendicular, and Frank and Nicholas (1953) refer to such dislocations ( $\mathbf{b}_3$ ) as doubtfully stable.

In addition to the perfect dislocations, we must consider the possibility of dissociation into allowable imperfect dislocations. The Burgers vectors of the possible imperfect dislocations are obtained by adding any lattice vector  $\mathbf{u}$  (including, of course,  $\mathbf{u} = 0$ ) to a non-lattice vector which will generate a low energy stacking fault. The energy of the imperfect dislocation is again supposed to be characterized by  $b^2$ , but in addition we must include the energy of the stacking fault which is associated with the imperfect dislocation.

We now consider the various common lattices and describe their stable dislocations and the important dislocation reactions. In the simple cubic lattice, the shortest lattice vectors joining nearest neighbours are of the form  $a\langle 100 \rangle$ , and dislocations with Burgers vectors equal to these are always stable. Vectors joining second and third nearest neighbours are of types  $a\langle 110 \rangle$  and  $a\langle 111 \rangle$  respectively, and the corresponding dislocations are of doubtful stability with respect to dislocations  $a\langle 100 \rangle$ . All other perfect dislocations are unstable, and there are no imperfect dislocations.

The shortest lattice vectors of the f.c.c. structure again join nearest atomic neighbours,

and are of the form  $\frac{1}{2}a\langle 110 \rangle$  when referred to the conventional unit cell; three non-coplanar lattice vectors of this type define the rhombohedral unit cell of the primitive lattice. The twelve dislocations of this type are stable with respect to other perfect dislocations. There are six second nearest neighbour vectors of type  $a\langle 100 \rangle$  and the corresponding dislocations are doubtfully stable; all other perfect dislocations are unstable.

The possible stacking faults in the f.c.c. lattice were described on p. 123. The operations involved in making intrinsic ( $1\Delta$ ) or extrinsic ( $2\Delta$ ) faults by inserting or removing close-packed planes correspond to displacements of  $(a/3)\langle 111 \rangle$  normal to these planes. Faults of both types may thus be bounded by imperfect dislocations having Burgers vectors equal to these displacements; as discussed previously, the sign of the Burgers vector is ambiguous until some convention is formulated. All possible imperfect dislocations of the f.c.c. structure are thus obtained by adding  $(a/3)\langle 111 \rangle$  to any lattice vector. Two faults of the same type may be bounded by different imperfect dislocations, just as different perfect dislocations may exist in the same lattice.

Frank and Nicholas (1953) have used a notation which gives a description of both dislocation and associated fault. They suppose the  $[111]$  direction to be positive upwards, and regard the left- and right-hand edges of the stacking fault as having the same direction, so that for a fault produced by adding or removing part of a plane these edges have opposite Burgers vectors. For ( $1\Delta$ ) faults, the two edges are called  $L$  and  $R$ , for ( $2\Delta$ ) faults  $\lambda$  and  $\rho$ , and the four imperfect dislocations are then

$$-\frac{a}{3}[111]L, \quad \frac{a}{3}[111]R, \quad \frac{a}{3}[111]\lambda, \quad -\frac{a}{3}[111]\rho.$$

The possible imperfect dislocations are now tested for stability against dissociation, as above. A dislocation, say of type  $L$ , may dissociate into a perfect dislocation and another  $L$  dislocation, or into two  $\lambda$  dislocations, and there are corresponding rules for the other types. This leads to the following  $L$  dislocations bounding ( $1\Delta$ ) stacking faults on a  $(111)$  plane.

$$\left. \begin{array}{l} (1) \quad -\frac{a}{3}[111]L, \\ (2) \quad \frac{a}{6}[11\bar{2}]L, \quad \frac{a}{6}[1\bar{2}1]L, \quad \frac{a}{6}[\bar{2}11]L, \\ (3) \quad \frac{a}{6}[411]L, \quad \frac{a}{6}[141]L, \quad \frac{a}{6}[114]L. \end{array} \right\} \quad (33.1)$$

Dislocations of types (1) and (2) are stable. Type (1) is the original sessile dislocation, first introduced by Frank (1949) and sometimes called a Frank partial. Type (2) has its Burgers vector in the plane of the stacking fault, which may be regarded as produced by slip on the  $(111)$  plane in a  $\langle 11\bar{2} \rangle$  direction (see deformation stacking faults, p. 124). These dislocations were introduced by Heidenreich and Shockley (1948) and are called half-dislocations or Shockley partials. Fig. 7.19 shows Frank and Shockley partials bounding

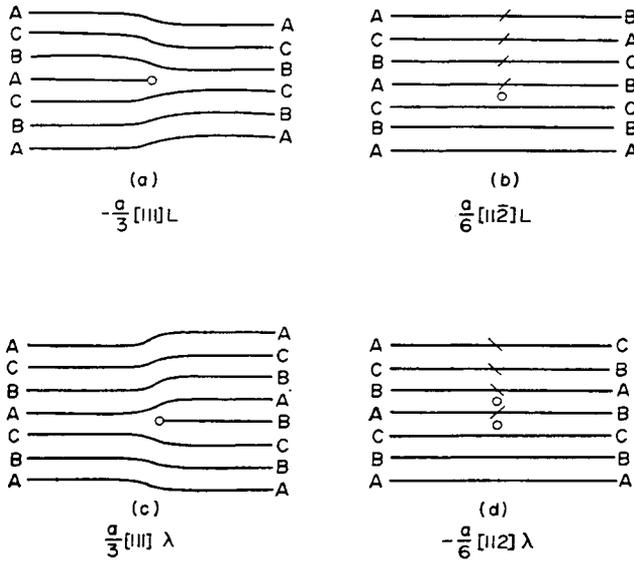


FIG. 7.19. Partial dislocations and stacking faults on a {111} plane of a f.c.c. structure (after Frank and Nicholas, 1953). o: Position of dislocation. Sloping lines indicate where changes must be made in the nomenclature of the planes.

the left-hand side of  $(1\Delta)$  and  $(2\Delta)$  faults respectively. Note that the  $(1\Delta)$  stacking fault is a one-layer twin on the tensile side of the sessile (edge) dislocation, whereas the  $(2\Delta)$  fault is a two-layer twin on the compressive side. Dislocations of type (3) are sessile; they are doubtfully stable against dissociations of type

$$\frac{a}{6} [411]L = \frac{a}{6} [2\bar{1}\bar{1}]\lambda + \frac{a}{3} [111]\lambda. \quad (33.2)$$

Since there are four sets of {111} planes, each of which may contain  $L$ ,  $R$ ,  $\lambda$ , and  $\varrho$  dislocations of the above types, there are in all 64 stable dislocations of types (1) and (2), and 48 doubtfully stable dislocations of type (3). Frank and Nicholas use the notation  $a$ ,  $b$ ,  $c$ ,  $d$  to denote the planes  $(1\bar{1}\bar{1})$ ,  $(\bar{1}\bar{1}1)$ ,  $(\bar{1}1\bar{1})$ , and  $(11\bar{1})$  respectively. Then any of the above imperfect dislocations is represented by the Burgers vector combined with a symbol such as  $D_i$ , where  $D$  stands for one of the quantities  $L$ ,  $R$ ,  $\lambda$ ,  $\varrho$ , and  $i$  for one of  $a$ ,  $b$ ,  $c$ ,  $d$ . The rules for obtaining all the dislocations from those of the  $L_d$  set described above are obvious.

The existence of imperfect dislocations may allow perfect dislocations to lower their energies by dissociation. Consider a dislocation  $\frac{1}{2}a[1\bar{1}0]$ . When it lies in a  $(111)$  plane, the dissociation

$$\frac{a}{2} [1\bar{1}0] = \frac{a}{6} [1\bar{2}1]L_d + \frac{a}{6} [2\bar{1}\bar{1}]R_d \quad (33.3)$$

is possible and leads to a decrease in energy. The two partial dislocations have parallel components in the  $[1\bar{1}0]$  direction, and thus repel each other. Since they are both glissile

in (111), they will separate, leaving an area of stacking fault, and forming an extended dislocation as described on p. 287. The separation does not continue indefinitely since the finite energy of the stacking fault supplies a constant attractive force, and the repulsive force decreases with the separation. When the forces balance, the total energy is minimized.

From eqn. (30.25) and the subsequent text, we see that the repulsive force between two parallel half-dislocations is of the form

$$f_r = \mu \mathbf{b}_1 \cdot \mathbf{b}_2 / 2\pi r \quad (33.4)$$

leaving out the small numerical factors which depend on the orientation of the dislocation line in relation to its Burgers vector. For half-dislocations,  $\mathbf{b}_1 \cdot \mathbf{b}_2 = a^2/12$ , so that  $f_r = \mu a^2 / 24\pi r$ . Heidenreich and Shockley (1948) calculated the equilibrium separation on the assumption that the stacking fault is effectively a hexagonal layer two atoms thick. The excess free energy of the hexagonal volume is then equivalent to a surface energy of the stacking fault of magnitude  $\sigma^f$  per unit area, so that  $\sigma^f$  is also the force per unit length of the bounding imperfect dislocations. Equating these forces, the equilibrium separation is

$$r = \mu a^2 / 24\pi \sigma^f. \quad (33.5)$$

In the approximation used by Heidenreich and Shockley, a free energy difference between the phases of  $420 \text{ J g atom}^{-1}$  corresponds to a surface energy of the fault of  $\sim 0.02 \text{ J m}^{-2}$  ( $20 \text{ erg cm}^{-2}$ ) and  $r$  is of the order of  $20a$ . The separation of the half-dislocations is thus small but much greater than the effective width of each. It is important to note that in this approximation the fault energy, and hence  $r$ , can be functions of temperature. In a metal which undergoes a transition from a f.c.c. to a h.c.p. phase, for example, the fault energy will become very small, and the partials will therefore tend to separate spontaneously as the transition temperature is approached. Actually, of course, the justification for using the macroscopic free energy difference between bulk phases in order to estimate the stacking fault energy is very slender, and it is rather surprising that in fact there seems to be a good empirical correlation between the occurrence of low fault energies and phase transitions between f.c.c. and h.c.p. phases. Another early method of estimating the fault energy is to regard it as a monolayer twin, so that the energy of the fault is equal to twice the experimental surface energy of a coherent twin interface. According to this approximation, intrinsic and extrinsic faults have the same energy. Estimates of this kind have been largely superseded by experimental methods which have been developed for measuring fault energies.

Seeger and Schoeck (1953) considered the formation of extended dislocations in anisotropic f.c.c. materials, making use of a method due to Leibfried and Dietze. Their results differ in important respects from those of the isotropic theory. In particular, when the extended dislocation has an edge orientation (line parallel to  $\langle 112 \rangle$ ) the separation is much greater than when it has a screw orientation (line parallel to  $\langle 110 \rangle$ ). For surface energies of 0.2, 0.04, and  $0.02 \text{ J m}^{-2}$  (then thought to be the appropriate energies for aluminium, copper, and cobalt respectively), the equilibrium separations in interatomic distances were estimated as 1.6, 12, and 50 for edge dislocations, and 1, 5, and 7 for screw dislocations. The effect of the high stacking fault energy is that the dislocations in aluminium are completely unextended, the calculated widths being smaller than the core radii.

The development of high resolution techniques such as weak-beam dark-field electron-microscopy (Cockayne *et al.*, 1969; Cockayne, 1972) has enabled the separation of the partial dislocations to be observed and measured with relatively high accuracy, even for metals such as silver and copper where the separations are small. Stobbs and Sworn (1971) and Cockayne *et al.* (1971) have measured the variation of the separation with orientation for copper, and the latter authors have obtained similar results for silver. In the case of copper, the observed spacing is 4 nm (40 Å) for edge orientations and 2 nm (20 Å) for screw orientations, and the edge spacing corresponds to a stacking fault energy of 0.041 J m<sup>-2</sup> (41 erg cm<sup>-2</sup>). There is very good agreement between the two sets of measurements, but it appears that the separations cannot be calculated solely from anisotropic elastic theory and an assumed constant value of  $\sigma^f$ , at least at the smaller separations. The ratio of the edge width to the screw width should be  $\sim 3.2$  for copper, but is observed experimentally to be  $\sim 2.1$ . Thus the effects of the dislocation cores and/or non-linear elastic interactions between the partials has to be included to explain the observed orientation dependence. This is illustrated also by calculations made by Cockayne *et al.* by the method of Seeger and Schoeck which, as mentioned above, is based on the Peierls model. This gives a stacking fault energy of 0.032 J m<sup>-2</sup>, which is more than 20% smaller than that obtained from anisotropic elasticity alone. The corresponding values obtained for silver are 0.016 and 0.014 J m<sup>-2</sup> on the basis of anisotropic theory and of the Seeger–Schoeck model respectively. The “best” values for the stacking fault energies of these metals are thus still a little uncertain, but the results now obtained are in reasonably good agreement with those obtained from the only other direct technique (measurement of the curvature of an extended dislocation line at a threefold node, see p. 304) and the spread compares very favourably with previous estimates by indirect techniques which, for copper, ranged up to 0.163 J m<sup>-2</sup> (Seeger *et al.*, 1959).

Instead of the dissociation represented by eqn. (33.3) it is also possible for the following reaction to occur:

$$\frac{a}{2} [1\bar{1}0] = \frac{a}{6} [1\bar{2}1] \rho_d + \frac{a}{6} [2\bar{1}\bar{1}] \lambda_d. \quad (33.6)$$

The energy of this pair of half-dislocations is expected to be slightly greater than that of the pair bounding the intrinsic fault, but it is possible that both types of dissociation occur. Loop annealing methods for comparing fault energies (see p. 305) suggest that for some metals at least, the energies of (1 $\Delta$ ) and (2 $\Delta$ ) faults are comparable, so that the equilibrium separations corresponding to equations (33.3) and (33.6) may be similar.

Each perfect dislocation of type  $\frac{1}{2}a\langle 110 \rangle$  may thus reduce its energy by extending in either of the two  $\{111\}$  planes which contain its Burgers vector. In a well-annealed f.c.c. crystal there may be a three-dimensional network of extended dislocations of this type, forming ribbons of stacking fault on all the close-packed planes, although, as we shall see, it is more common for most of the dislocations to be concentrated into planar nets. Whenever an extended dislocation turns out of a particular  $\{111\}$  plane, a node may be formed and the partial dislocations come together, but this need not happen (see below).

When a  $\frac{1}{2}a\langle 110 \rangle$  dislocation lies in a  $\{111\}$  plane which does not contain its Burgers

vector, it cannot glide. Nevertheless a dissociation may occur; for example in the plane  $(1\bar{1}\bar{1})$ , a  $\frac{1}{2}a[1\bar{1}0]$  dislocation may split as follows

$$\frac{a}{2}[1\bar{1}0] = \frac{a}{6}[1\bar{1}2]L_a + \frac{a}{3}[1\bar{1}\bar{1}]R_a. \quad (33.7)$$

This reaction, first considered by Cottrell and Bilby (1951), produces no first-order change in elastic energy, and thus occurs only under the influence of stress, since the energy of the stacking fault must be supplied externally. The  $R_a$  dislocation is sessile and remains in the site of the original dislocation, but the  $L_a$  dislocation may glide away, leaving a region of  $(1\Delta)$  stacking fault. Once again, the analogous dissociation into  $\lambda_a$  and  $\rho_a$  dislocations is also possible. We thus see that each  $\frac{1}{2}a\langle 110 \rangle$  dislocation may dissociate in eight ways, four of them in the two  $\{111\}$  planes in which it can glide, and four in the other two  $\{111\}$  planes.

Kuhlmann-Wilsdorf (1958) has pointed out that in appropriate circumstances, we may imagine the reaction of eqn. (33.7) to proceed from right to left. Consider the formation of a dislocation by the collapse of a disc of vacancies aggregated on a  $\{111\}$  plane in the manner described on p. 244. As already discussed, this will produce the  $R_a$  dislocation of eqn. (33.7) as a circular loop enclosing a disc of stacking fault. However, if the stacking fault energy is sufficiently high, the lattice might not tolerate this situation, and a loop of Shockley partial may be spontaneously nucleated to combine with the Frank partial and give the resultant perfect dislocation loop. The vacancies have then condensed to give a prismatic loop of lattice dislocation, which has a higher strain energy than the Frank partial it has replaced, but a lower total energy because of the elimination of the stacking fault.

Calculations show that it is energetically favourable to form a faulted loop rather than an unfaulted loop on initial collapse of a small vacancy disc, even for quite high values of  $\sigma^f$ . A faulted loop may continue to grow by climb if there is a vacancy supersaturation, and it will eventually become metastable relative to an unfaulted loop. However, the Shockley partial required to remove the fault must now be nucleated on the periphery of the fault, and the activation energy for this event is very large ( $\sim 10$  eV), so that once a faulted loop has formed it is difficult for it to be converted (Saada, 1962; Saada and Washburn, 1963). In metals quenched to produce high vacancy supersaturations, both faulted and unfaulted loops have been observed under different circumstances; faulted loops are observed in aluminium, for example, only with very pure specimens quenched from relatively low temperatures. High temperatures or impurities were formerly believed to assist the thermal nucleation of Shockley partials, but it appears that the transition depends mainly on stress-assisted nucleation (Edington and Smallman, 1965).

The imperfect dislocations we have considered have all separated stacking faults from good crystal. It is also possible to have two different stacking faults meeting on the same plane, and the line of separation is then another type of imperfect dislocation. To define its Burgers vector we make two Burgers circuits, one beginning on the first fault and one on the second fault, and take the difference of the two closure failures. These dislocations are thus combinations of the simple imperfect dislocations already described; from the nature

of the definition, they result from the combination of the right-hand side of one fault with the left-hand side of the other, and are thus of the form  $L_d \rho_i$  or  $R_d \lambda_i$ . A typical dislocation of type  $L_d \rho_d$  is  $(a/3)$  [111], and all the Burgers vectors of dislocations of this group are the same as those of the  $R_d$  group. Equally all those of the  $R_d \lambda_d$  group are equivalent to  $L_d$  vectors. Corresponding to the existence of these imperfect dislocations, we may have more complex extended dislocations in which translation twin surfaces of different types are successively joined together. The total Burgers vectors of these extended dislocations are rather large.

Two non-parallel  $\{111\}$  planes meet in a  $\langle 110 \rangle$  direction. If each of the planes contains a stacking fault, the line of intersection of the two stacking faults will be an imperfect dislocation. This imperfect dislocation may be obtained by combining two previous imperfect dislocations lying on separate planes; there is a threefold node of imperfect dislocation lines at each end of the line of intersection of the stacking faults. Thompson (1953), following a suggestion of Nabarro, has called these stacking fault intersections "stair-rod dislocations". Frank and Nicholas show that there are in fact 96 different sets of the form  $D_i D_j'$ , where  $i$  and  $j$  are different and  $D$  and  $D'$  may represent the same or different symbols. However, all of these dislocations may be derived from the simple forms  $L_a L_b$  ( $= -R_a R_b$ ) in which the included angle between the two stacking faults is obtuse and  $L_a R_b$  ( $= -R_a L_b$ ) in which the included angle is acute. The vector  $(a/3)$  [00 $\bar{1}$ ]  $L_a L_b$  gives the only stable type of dislocation of the first group. In the second group, there are stable Burgers vectors of type  $(a/6)$  [ $\bar{1}$ 10]  $L_a R_b$  and  $(a/3)$  [ $\bar{1}$ 10]  $L_a R_b$ ; all other possible Burgers vectors are unstable. Frank and Nicholas give a table for deriving the correct representation of any dislocation from the type vectors of the two basic groups.

When two extended dislocations on intersecting planes meet in a stair-rod dislocation, a reduction in energy is possible if the partials bounding the stacking faults curve inwards towards the line of intersection. The increase in energy due to the increased interactions of the partial dislocations is then compensated by the reduction in length of the stair-rod dislocation. In the limit, there is zero length of stair-rod dislocation, each extended dislocation being fully constricted to give a fourfold node or partial dislocations. This situation was once assumed to exist in all circumstances, but is really a limiting case and need not give the lowest energy. A related problem is the structure of a jog in an extended dislocation, which may or may not be fully constricted. According to Hirsch (1962) the jog configurations may be analysed into rather complex arrangements of partial dislocations.

Stair-rod dislocations are necessarily edge and sessile. Since they exist only at the intersection of two stacking faults, they are in fact "supersessile", and can neither glide nor climb without first dissociating into component partials. However, if both stacking faults arise from extended dislocations which move in their respective glide planes, it is possible that the stair-rod dislocation is moved along parallel to its length. What really happens in such motion is that the leading two partials (one on each plane) create additional length of stair-rod as they glide, and the trailing partials remove some of the existing stair-rod.

Most attention has been paid to the stair-rod dislocation of type  $(a/6)$  [ $\bar{1}$ 10]  $L_a R_b$ . This was first described by Cottrell (1952) following a dislocation reaction suggested by Lomer (1951). Lomer supposed two perfect dislocations,  $\frac{1}{2}a$ [10 $\bar{1}$ ] and  $\frac{1}{2}a$ [011], the first lying in the

(111) or  $d$  plane and the second in the ( $\bar{1}\bar{1}1$ ) or  $c$  plane. Then the following reaction leads to a reduction in energy

$$\frac{a}{2} [10\bar{1}] + \frac{a}{2} [011] = \frac{a}{2} [110]. \quad (33.8)$$

The  $\frac{1}{2}a[110]$  perfect dislocation is an ordinary stable dislocation of the structure, but it must lie along a  $[1\bar{1}0]$  line which is the intersection of the two glide planes. The dislocation is thus an edge, and can glide only in the plane (001) which is not a usual glide plane in the f.c.c. structure. Cottrell pointed out that the energy can be further reduced by the dissociation

$$\frac{a}{2} [110] = \frac{a}{6} [11\bar{2}] + \frac{a}{6} [112] + \frac{a}{6} [110]. \quad (33.9)$$

The  $(a/6)\langle 112 \rangle$  dislocations move away from the  $[1\bar{1}0]$  line in the two glide planes, leaving stacking faults in these planes, and the stair-rod type dislocation  $(a/6)[110]$  is left at the line of intersection. Clearly, the same result is obtained by supposing the original dislocations to be dissociated into Shockley partials, so that in the extended notation

$$\frac{a}{2} [10\bar{1}] = \frac{a}{6} [11\bar{2}]L_d + \frac{a}{6} [2\bar{1}\bar{1}]R_d,$$

$$\frac{a}{2} [011] = \frac{a}{6} [112]R_c + \frac{a}{6} [\bar{1}21]L_c,$$

and the reaction is then

$$\frac{a}{6} [2\bar{1}\bar{1}]R_d + \frac{a}{6} [\bar{1}21]L_c = \frac{a}{6} [110]L_cR_d. \quad (33.10)$$

The group of three partial dislocations is very important in some theories of work hardening, which assume that Lomer–Cottrell barriers are stable obstacles to glide dislocations. However, calculations by Stroh (1956) indicate that even an infinite barrier of this type is much weaker than once believed, and will support only a small number of dislocations at the stress levels attained in plastic deformation. A real barrier is expected to be finite in extent, ending at nodes, and this should give way by “unzipping” from the ends. Hence many of the more recent work-hardening theories do not attribute any very special significance to Lomer–Cottrell barriers. Experimental observations in the electron-microscope have revealed the early formation of dislocation networks by interactions of the kind discussed above, but although Lomer–Cottrell barriers are observed (e.g. Basinski, 1964), similar interactions between glide and forest dislocations which do not lead to sessile configurations seem to be equally important.

The above description does not quite exhaust the imperfect dislocations of the f.c.c. structure. It is possible to have three stacking faults meeting in a line, two (of intrinsic and extrinsic types respectively) lying in one plane and the third in an intersecting plane. Similarly, if two intersecting planes each have two types of fault, all four may meet in the line of intersection. In the first case, we get dislocation combinations of types  $L_a\varrho_aL_b$  and  $L_a\varrho_aR_b$ , and these have vectors identical with those of  $R_aL_b$  and  $R_aR_b$  respectively. Moreover,

the stable dislocations have Burgers vectors identical with those of the stable stair-rod dislocations. The four dislocation combinations, e.g.  $L_a \rho_a L_b \rho_b$  or  $L_a \rho_a R_b \lambda_b$ , have Burgers vectors corresponding to those of the stable  $R_a R_b$  and  $R_a L_b$  sets.

The notation used in the above description of the f.c.c. structure is rather clumsy because of the necessity of specifying the type of stacking fault as well as the Burgers vector of the partial. If the stacking fault specification is omitted, or if these are assumed to be always of the  $(1\Delta)$  type, a considerable simplification is possible, especially if a geometrical representation due to Thompson (1953) is used. The four sets of  $\{111\}$  planes correspond to the faces of a regular tetrahedron, shown opened out in Fig. 7.20. The vertices opposite

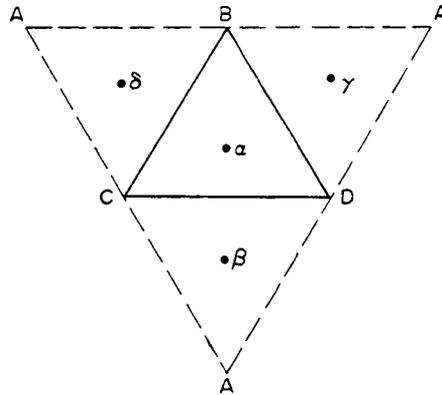


FIG. 7.20. The Thompson tetrahedron.

to the faces  $a, b, c, d$  are marked  $A, B, C, D$ , and the mid-points of these faces by  $\alpha, \beta, \gamma, \delta$ . The stable dislocations of the f.c.c. structure then have Burgers vectors with the following representations:

- (1) Perfect dislocations of type  $\frac{1}{2}a\langle 110 \rangle$  are given by the edges  $AB$ , etc., the letters and their order covering the twelve cases.
- (2) Perfect dislocations (doubtfully stable) of type  $a\langle 100 \rangle$  are represented by symbols  $AB/CD$  and permutations in which no attention is paid to the order of the grouped letters (i.e.  $AB/CD = BA/CD = -CD/AB$ ). The symbol means a vector equal to twice the join of the mid-points of  $AB$  and  $CD$ , and this is equal to the vector sum of  $AC$  and  $BD$ .
- (3) Imperfect dislocations of type  $(a/3)\langle 111 \rangle$  are represented by lines  $\alpha A$ , etc. No distinction is possible between  $L$  and  $\lambda$  dislocations. Dislocations of type  $R(\rho)$  have the same symbols as the  $L(\lambda)$  dislocations with the order of the letters reversed.
- (4) Imperfect dislocations of type  $(a/6)\langle 211 \rangle$  are represented by symbols  $\alpha B, \alpha C$ , etc.
- (5) Doubtfully stable imperfect dislocations of type  $(a/6)\langle 411 \rangle$  are represented by  $D/AB$  and permutations.
- (6) Stair-rod dislocations of type  $L_a L_b$  are given by symbols  $\alpha\beta/CD$  etc.
- (7) Stair-rod dislocations of type  $(a/6)\langle 110 \rangle L_a R_b$  are given by  $\alpha\beta$ , etc., and those of types  $(a/3)\langle 110 \rangle L_a R_b$  by  $A\alpha/B\beta$ , etc.

Thompson's notation has considerable attractions, since the geometrical relations between the Burgers vectors and the close-packed planes are readily visualized, and it has been used in the literature much more than that of Frank and Nicholas. Only when the possibility of extrinsic stacking faults is being specifically considered will it be necessary to use the more elaborate method. As an example of the simplicity of the tetrahedron representation, consider eqn. (33.7). This becomes

$$BA = B\alpha + \alpha A \quad (33.11)$$

Another example arises from the dislocation reaction

$$\frac{1}{2}a[1\bar{1}0] + \frac{1}{2}a[01\bar{1}] = \frac{1}{2}a[10\bar{1}] \quad (33.12)$$

in which two dislocations with non-parallel Burgers vectors in the same (111) slip plane combine to form a new dislocation with corresponding reduction in energy. If the reaction occurs over a limited length, a node is formed and, in terms of the Thompson tetrahedron, (33.12) may be written more symmetrically as

$$BA + AC + CB = 0. \dagger \quad (33.13)$$

Nodes of this kind occur in networks, and are very stable when the dislocation lines are mutually at  $120^\circ$ . By means of the Thompson tetrahedron it may now be shown that if the dislocations are all extended according to (33.11), the nodes in a hexagonal network are alternately extended and constricted provided only that all faults are of the same type. This is, indeed, observed in many metals, and following a suggestion by Whelan (1959) measurements of the radius of curvature at an extended node have been used by several authors to estimate the stacking fault energy. The method depends on the equilibrium between the mutual and self-stresses of the dislocations (see p. 267) and the tension due to the fault, and is now capable of giving reasonable accurate results for materials with moderately low fault energies (see Gallagher, 1968, and Gallagher and Liu, 1969, for experimental values).

If the extrinsic fault energy is sufficiently low, all threefold nodes are extended, but enclose stacking faults which are alternately of  $1\Delta$  and  $2\Delta$  types with cross-over points (two-fold constrictions) between the nodes. In principle it is now possible to determine the ratio of the energies of the two types of fault although there are difficulties caused by the energy of the twofold constrictions. Gallagher (1966) found the ratio of the extrinsic to the intrinsic fault energies in two silver-indium alloys to be  $\sim 1.09$  and  $\sim 1.03$ .

Yet another example of the utility of the Thompson tetrahedron is provided by an analysis of the possible structures of long jogs in extended dislocations and of related defects such as extended dipoles. If a jog lowers its energy by dissociating in a plane inclined to the slip planes of the extended dislocations, the resulting configurations (Thompson, 1955; Hirsch, 1962) can be quite complex. The motion of dislocations containing dissociated long jogs can lead to the trailing of faulted dislocation dipoles (i.e. long strips of stacking fault on an inclined  $\{111\}$  plane, bounded by Frank partials), and in Hirsch's theory these dipole loops are always of vacancy kind since jogs which would trail interstitial-type dipoles either

move conservatively or leave unfaulted dipoles. This asymmetry disappears if the dissociation allows  $2\Delta$  faults to form as readily as  $1\Delta$  faults. Faulted dipoles can also occur by the dissociation of unfaulted dipoles, obtained, for example, from interactions of dislocations from different sources. Seeger (1964) pointed out that further dissociation of a faulted dipole in a  $\{111\}$  plane could occur by dissociation of the opposite Frank partials into stair-rods and Shockley partials, the latter dislocations moving off in parallel slip planes to give a final configuration which is Z-shaped in section if all three faults are of  $1\Delta$  type, and S-shaped if the middle fault is  $2\Delta$ . Faulted dipoles appear in electron-micrographs as thin straight lines, and careful contrast experiments are needed to distinguish them from similarly oriented Lomer-Cottrell dislocations. Seeger proposed that observations on these defects would provide a method for measurement of stacking fault energies and following developments by several authors, Steeds (1967) formulated an anisotropic elastic theory of the interactions, and a detailed theory of the fault contrast. He used his experimental measurements on copper, silver, and gold to estimate values of the fault energies of these materials, but the values obtained are rather high in comparison with those given by other direct methods.

Elementary steps between stacking faults on adjacent close-packed planes will also have complex configurations when the dislocations are extended, and are sometimes known as jog-lines (Thompson, 1955). Jog-lines are dipole-like defects corresponding to the overlapping of the cores of two partial dislocations, and there are four basic kinds which are equivalent geometrically to lines of one-third or two-thirds vacancies or one-third or two-thirds interstitials. The lowest energy configuration is the one-third vacancy type, which corresponds formally to a  $\gamma\delta + \delta\gamma$  dipole. Hirsch applied arguments similar to those applied to long jogs to discuss the motion of dislocations containing elementary jogs, and used this as the basis of a theory of work-hardening.

We have already referred to planar Frank loops formed by vacancy condensation, but more complex loops which involve stepped stacking faults have also been detected (Clarebrough *et al.*, 1966; Morton and Clarebrough, 1969). Although some of these could arise from two independently nucleated planar loops, some observations require the concept of the climb of the fault, as first considered by Escaig (1963) and Schapink and de Jong (1964). These authors showed that a row of four vacancies absorbed on a fault may be changed by a simple atomic shift into a triangular array of one-third vacancy jog-lines. Subsequent vacancies can then be absorbed one by one, each vacancy leading to an increase in the length of the triangular fault by one atom distance. This process progressively transfers the stacking fault on to the adjacent plane, and is more probable in materials of very low fault energy where addition of partials at the edge of a growing fault (i.e. climb of the Frank partial rather than of the fault) may be inhibited because of the dissociation (33.14) which is considered below.

A different situation is the formation of multiple loops by condensation of vacancies on successive  $\{111\}$  planes, first observed by Westmacott *et al.* (1961). In later work, two-, three- and four-layer defects have been observed (see Edington and West, 1966, 1967) in quenched aluminium and aluminium alloys. The observations allow the deduction that the extrinsic fault energy is less than twice the intrinsic energy; the observation of four-

layer defects, however, suggests that heterogeneous nucleation on some impurity has occurred.

In many quenched f.c.c. metals of low fault energy, a characteristic defect observed in the electron-microscope is a more or less regular tetrahedron of stacking fault, the edges of the tetrahedron being stair-rod dislocations. The Thompson model provides a particularly simple way of discussing this fault.

Suppose a flat disc of vacancies collapses to give a Frank sessile dislocation which has a triangular shape, the edges being parallel to the  $\langle 110 \rangle$  directions in the  $\{111\}$  plane of the disc. If the stacking fault energy is sufficiently high, the Kuhlmann-Wilsdorf reaction (the reverse of eqn. (33.7)) will take place to give a prismatic loop of perfect dislocation, eliminating the fault at the expense of some elastic energy. If the fault energy is sufficiently low, however, there will be a reverse tendency to lower the elastic energy at the expense of some increase in the area of fault. The Frank partial lying along a close-packed direction may dissociate into a low-energy stair-rod dislocation and a Shockley partial dislocation. The Shockley partial cannot glide in the plane of the original vacancy disc, but it can glide in the plane which intersects this in the direction of the Frank partial considered. A typical such reaction in the  $d$  plane, for example, is

$$\frac{a}{3} [111] = \frac{a}{6} [101] + \frac{a}{6} [121], \quad (33.14)$$

or, in terms of the Thompson model,

$$\delta D = \delta\beta + \beta D. \quad (33.15)$$

The  $\beta D = (a/6) [121]$  dislocation can glide in the  $b$  plane, leaving the  $\delta\beta$  stair-rod dislocation at the intersection of the two stacking faults which meet at an acute angle. This reaction is favourable so far as the elastic energy of the partial dislocations is concerned.

Now consider three such dissociations to take place, corresponding to the three orientations of the triangular Frank partial. Using a partial in the  $a$  plane, these would be

$$\left. \begin{aligned} \alpha A &= \alpha\beta + \beta A, \\ \alpha A &= \alpha\gamma + \gamma A, \\ \alpha A &= \alpha\delta + \delta A. \end{aligned} \right\} \quad (33.16)$$

The Shockley partials will begin to bow out on the three  $\{111\}$  planes which intersect the  $a$  plane along the sides of the original Frank dislocation. If we use the sign convention suggested on p. 245, looking outwards from each vertex of the original triangle, the Shockley partials will have opposite signs at their two ends. These partials attract one another in pairs, according to equations of type

$$\left. \begin{aligned} \beta A + A\gamma &= \beta\gamma, \\ \gamma A + A\delta &= \gamma\delta, \\ \delta A + A\beta &= \delta\beta, \end{aligned} \right\} \quad (33.17)$$

so that the four lines of dislocation radiating from each corner of the triangle become three lines, and the nodes where the pairs of Shockley partials meet move away from the original

a plane until they all join together in a single nodal point. The reactions (33.17) are all of the type (33.10), and hence also lead to a reduction of energy. The final result is the defect we described above, a tetrahedron of stacking fault with stair-rod dislocations as its edges. The tetrahedron may be identified with the Thompson tetrahedron, and the edges  $AB$ ,  $BC$ ,  $CD$ ,  $DA$  have Burgers vectors  $\gamma\delta$ ,  $\delta\alpha$ ,  $\alpha\beta$ ,  $\beta\gamma$  which form a tetrahedron inverse to that of the dislocation lines themselves. The above theory is due to Silcox and Hirsch (1959) who also made the first observations of these defects in gold. Clearly the final defect is highly symmetrical, and an alternative to the two-stage nucleation process just described is the direct collapse of small three-dimensional vacancy clusters into tetrahedra which subsequently grow by further vacancy addition (de Jong and Koehler, 1963). Both these processes probably occur in quenched metals but, in addition, tetrahedra appear to form by a dislocation reaction during deformation (Loretto *et al.*, 1965).

The final defect consists of six partial dislocation lines, the elastic energy of each of which is approximately proportional to  $(a^2/18)$  per unit length, whereas the equivalent loop of Frank partial consists of three dislocation lines of the same length, and elastic energy of  $(a^2/3)$  per unit length. The elastic energy of the tetrahedral defect is thus only about one-third of that of the corresponding loop of Frank partial formed by an equivalent number of vacancies condensing. On the other hand, the finite stacking fault energy sets an upper limit to the size of the tetrahedral regions which may be formed; an approximate calculation suggests that with a fault energy of  $0.033 \text{ J m}^{-2}$ , the edge of the tetrahedron will be about 43 nm.

Under conditions of vacancy supersaturation, tetrahedra increase in size by further vacancy condensation. Each face could climb by the nucleation of a high-energy jog-line along the edge of the tetrahedron or by the formation of a low-energy (one-third vacancy) jog-line at the apex; the latter process is much more probable (Kuhlmann-Wilsdorf, 1965). Shrinkage of a tetrahedron by vacancy emission is more difficult to initiate, since the step nucleus is either a short length of high-energy jog-line at the apex, or a long low-energy jog-line. In practice, tetrahedra are found to be very stable against dissolution, and persist in gold up to  $\sim 800^\circ\text{C}$ .

Since triangular Frank faults and tetrahedra are sometimes observed in the same specimens, considerable attention has been paid to the conditions governing the transition. Simple theory, as outlined above, shows that the tetrahedron has approximately four times the fault energy and one-third the elastic self-energy of the equivalent Frank loop, so that as its size increases the tetrahedron eventually becomes less stable than the loop. However, a full calculation must include the elastic interaction energies of all the partials and is difficult even in isotropic approximation. The transition from tetrahedron to planar loop essentially involves the inverse of the Silcox-Hirsch mechanism, and at an intermediate stage the defect has the form of a truncated tetrahedron. After various calculations of the energy of this defect, Humble and Forewood (1968a, b) concluded that there is essentially no energy barrier to the transition and that experimental measurements of the largest tetrahedron and smallest Frank loop in a plastically deformed material give a reliable indication of the stacking fault energy.

The kinetics of growth or shrinkage of prismatic loops, faulted loops, or tetrahedra

may be followed by *in situ* experiments in the electron-microscope. Most of these observations have been made on aluminium, and it appears that the rate-limiting process is usually point defect diffusion to or from the foil surface (Seidman and Balluffi, 1966; Dobson *et al.*, 1967). Faulted loops anneal at a different rate from unfaulted loops because of the dominating effect of the fault energy on the effective climb force. Estimates of fault energy may be made from the climb rates but are subject to some uncertainty.

The aggregation of point defects in quenched or irradiated f.c.c. metals is intrinsically important, but also has great relevance, especially in aluminium alloys, to precipitation and age-hardening processes. Although the existence of vacancy-solute interactions introduces some important modifications into the processes described above, we defer further consideration to Part II, Chapter 16. It remains finally to consider whether any dislocations in the f.c.c. structure are unstable in particular orientations.

Since the Peierls force is believed to be small for f.c.c. metals any appreciable variation in  ${}^LW_d$  with  $\varphi$  must come from  ${}^LW_s$ . The calculations made by Head show that elastic instability for lattice dislocations arises only when both the elastic factors mentioned on p. 295 are rather large, and this excludes most common metals, but regions of instability are predicted for elastically anisotropic alloys such as indium-thallium. The individual Shockley partials of an extended dislocation are unstable under less rigid conditions; for example, over a range  $\varphi \approx 79\text{--}101^\circ$  for copper. Experimental observations on copper-aluminium alloys show that one of the partials of an extended dislocation adopts a zigzag configuration as predicted by theory (Clarebrough and Head, 1969).

We now turn to consider the h.c.p. structure, which is the only common metallic double lattice structure. This means that not all nearest neighbour atomic translations are possible Burgers vectors of perfect dislocations. There are, in fact, six stable perfect dislocations of type  $(a/3)\langle 11\bar{2}0 \rangle$  which represent nearest neighbour translations within the close-packed planes. Perfect dislocations of type  $c\langle 0001 \rangle$  are also stable, this being the next smallest lattice vector. Finally, there are possible Burgers vectors of type  $\left\langle \frac{a}{3} \frac{a}{3} \frac{2a}{3} c \right\rangle$ , but these dislocations are doubtfully stable against dissociation into one each of the above two types.

Vectors in the h.c.p. structure have been written above (see p. 39) in terms of the conventional four-axis reference system, which is adopted to ensure that equivalent planes and directions have similar indices. This advantage, however, is much reduced by the difficulty of handling four indices, and it is often more convenient to refer vectors to the three axes of the conventional h.c.p. cell, or of the orthohexagonal cell. With the conventional three-axis system, the above dislocations have Burgers vectors  $a\langle 100 \rangle$  and  $a\langle 110 \rangle$  for the first type,  $c\langle 001 \rangle$  for the second, and  $\langle a0c \rangle$  and  $\langle aac \rangle$  for the third type.

The possible low-energy stacking faults of the h.c.p. structure were discussed in Section 17, and it remains to enumerate the Burgers vectors of the dislocations which bound them. A typical displacement producing a  $(1\Delta)$  fault is  $\left[ \frac{2a}{3} \frac{a}{3} \frac{c}{2} \right]$  (expressed in the three-axis system), i.e. a nearest neighbour displacement between atoms not in the same basal plane. With a suitable choice of axes, the Burgers vector of either the left-hand side of a  $(1\Delta)$

fault or the right-hand side of a  $(1\bar{\nabla})$  fault is given by this displacement, and there are five other equivalent vectors. The  $L(1\Delta)$  or  $R(1\bar{\nabla})$  set may thus be

$$\left[ \frac{2a}{3} \frac{a}{3} \pm \frac{c}{2} \right], \quad \left[ -\frac{a}{3} - \frac{2a}{3} \pm \frac{c}{2} \right], \quad \left[ -\frac{a}{3} \frac{a}{3} \pm \frac{c}{2} \right], \quad (33.18)$$

and the negative of these give the  $L(1\bar{\nabla})$  or  $R(1\Delta)$  sets. As mentioned previously, the distinction between  $\Delta$  and  $\bar{\nabla}$  depends only on the choice of axes in the basal plane, and the two sets are interchanged by a rotation of  $60^\circ$ .

The deformation fault  $(2\Delta)$  corresponds to slip on the basal planes, and the dislocations which bound it thus correspond to the half-dislocations of the f.c.c. structure. With the above axes, the  $L(2\Delta)$  and the  $R(2\bar{\nabla})$  dislocations have Burgers vectors

$$\left[ \frac{a}{3} - \frac{a}{3} 0 \right], \quad \left[ \frac{a}{3} \frac{2a}{3} 0 \right], \quad \left[ -\frac{2a}{3} - \frac{a}{3} 0 \right], \quad (33.19)$$

and the  $L(2\bar{\nabla})$  and the  $R(2\Delta)$  dislocations have opposite Burgers vectors.

The extrinsic fault  $(3\Delta)$  may be produced by a displacement of  $c/2$  perpendicular to the close-packed planes, and the corresponding dislocations are thus rather similar to sessile dislocations of the f.c.c. structure. The two possible Burgers vectors  $[0 0 \mp \frac{1}{2}c]$  may be associated with either edge of  $(3\Delta)$  or  $(3\bar{\nabla})$  faults.

Application of the stability rule shows that the dislocations bounding  $(2\Delta)$  and  $(3\Delta)$  faults are stable, but that those bounding  $(1\Delta)$  faults are only doubtfully stable against dissociations of the type

$$\left[ \frac{2a}{3} \frac{a}{3} \frac{c}{2} \right] L(1\Delta) = \left[ \frac{2a}{3} \frac{a}{3} 0 \right] L(2\Delta) + \left[ 0 0 \frac{c}{2} \right] L(3\Delta). \quad (33.20)$$

The dislocations bounding  $(1\Delta)$  faults and  $(3\Delta)$  faults are sessile, but those bounding  $(2\Delta)$  faults are glissile, so that the above reaction could occur if energetically favoured. In addition to the vectors listed above, there are three sets of imperfect dislocations which may bound  $(2\Delta)$  faults, but all of these are doubtfully stable against dissociation into a perfect dislocation and a  $(2\Delta)$  dislocation of the above type.

The existence of low energy-faults allows stable perfect dislocations which lie in the close-packed planes to lower their energies by dissociation into imperfect dislocations. The reaction analogous to eqn. (33.3) is

$$a[100] = \frac{a}{3} [1\bar{1}0] L(2\Delta) + \frac{a}{3} [210] R(2\Delta). \quad (33.21)$$

Since the  $(2\Delta)$  dislocations are glissile, the reaction will occur spontaneously in the basal planes, and the equilibrium separation will depend on the stacking fault energy, as in the f.c.c. case. Calculations by Seeger (1955b) suggest that for zinc, cadmium, and magnesium, the stacking fault energy is high and the width of the extended dislocations less than  $0.7 \text{ nm}$  ( $7 \text{ \AA}$ ). On the other hand, h.c.p. cobalt has a very low stacking fault energy.

The other types of perfect dislocation may also reduce their energies by dissociation in the basal planes, e.g.

$$\left. \begin{aligned} c[001] &= \left[ \frac{2a}{3} \frac{a}{3} \frac{c}{2} \right] L(1\Delta) + \left[ -\frac{2a}{3} - \frac{a}{3} \frac{c}{2} \right] R(1\Delta), \\ [a0c] &= \left[ \frac{2a}{3} \frac{a}{3} \frac{c}{2} \right] L(1\Delta) + \left[ \frac{a}{3} - \frac{a}{3} \frac{c}{2} \right] R(1\Delta). \end{aligned} \right\} \quad (33.22)$$

The ( $1\Delta$ ) dislocations are sessile, and these separations may thus be achieved only by diffusive processes (dislocation climb), and are much less probable. The extended dislocation would, of course, be sessile.

As with f.c.c. structures, there is no limit to the number of geometrically possible extended dislocations, formed by joining together stacking faults of different types to give rather large resultant Burgers vectors. However, if only basal plane faults are possible, all faults are parallel in this structure, so there is no equivalent to the f.c.c. stair-rod dislocations. In some h.c.p. metals in which the axial ratio is close to the ideal value for close-packing, or is less than this, the most prominent slip planes may be prismatic  $\{10\bar{1}0\}$  or pyramidal  $\{10\bar{1}1\}$  planes. Atomistic calculations with Morse potentials (Schwartzkopff, 1969) show that stacking faults on these planes may be metastable. However, the computed energies are very high, and there is no experimental evidence to support the hypothesis.

The collapse of a monolayer disc of vacancies on a basal plane may give either a  $3\Delta$  fault with a low-energy partial dislocation, or a  $1\Delta$  fault with a higher-energy partial dislocation. This is analogous to the production of a faulted loop with a Frank partial or an unfaulted loop with a prismatic dislocation in a f.c.c. material, and provided the  $3\Delta$  fault does indeed have the higher energy, the  $1\Delta$  loop should be stable above some critical radius. Conversion of the  $3\Delta$  to a  $1\Delta$  fault by the inverse of the dissociation (33.20) requires the nucleation of a "Shockley partial" (as in the analogous f.c.c. reaction), but it is probable that in many metals vacancies condense directly to  $1\Delta$  loops (e.g. Lally and Partridge, 1966).

Precipitation of vacancies on two successive (001) planes will give unfaulted prismatic loops of types  $[00c]$  or  $[a0c]$ . In principle loops of the latter type may lower their energies by dissociation into  $[00c]$  loops +  $[a00]$  loops which then collapse. Also both loops may dissociate by climb according to (33.22); at an intermediate stage the  $1\Delta$  fault is contained between two concentric loops, the inner of which is shrinking and the outer growing by diffusion across the annulus until a single faulted loop enclosing twice the area of the original loop is obtained. In conditions of high vacancy supersaturation the reverse process has been observed, so that a second layer of vacancies nucleates on a faulted loop, and both of the concentric dislocation loops then grow outwards. Measurements of the kinetics of loop climb in foils have been used to estimate fault energies in some hexagonal metals by methods similar to those used for f.c.c. metals (Harris and Masters, 1966a, b; Dobson and Smallmann, 1966; Hales *et al.*, 1968).

Finally we consider the b.c.c. structure. Using the axes of the conventional unit cell as the reference system, we find there are two distinct types of stable perfect dislocation. There are eight vectors of type  $\frac{1}{2}a\langle 111 \rangle$  which define the displacements from a lattice point to its

nearest neighbour lattice points, and six vectors of type  $a\langle 100 \rangle$  which represent the next nearest neighbour displacements. All the other possible perfect dislocations are unstable. Perfect dislocations of type  $\frac{1}{2}a\langle 111 \rangle$  gliding on intersecting  $\{110\}$  slip planes will attract each other and form an edge type  $a\langle 100 \rangle$  dislocation with reduction in elastic energy. This is one suggested mechanism for crack nucleation in b.c.c. structures, which are usually brittle at low temperatures, since the width of the  $\langle 100 \rangle$  dislocation may be so narrow that it can be regarded as an incipient crack (Cottrell, 1958). There is some evidence for crack nucleation from intersecting slip bands in non-metallic materials such as magnesium oxide, but many b.c.c. brittle fractures begin from twin intersections (Hull, 1960).

As explained on p. 127, there is no firm evidence for the existence of monolayer stacking faults and partial dislocations in b.c.c. metals except under anomalous conditions, and theoretical calculations lead to the conclusion that single-layer faults are not mechanically stable on either  $\{112\}$  or  $\{110\}$  planes. Prior to this work, there was much speculation about possible dissociations of lattice dislocations with the production of faults on these planes; for example, a monolayer twin fault on  $(112)$  could have a left-hand edge bounded by a partial dislocation with Burgers vector either  $[-a/6, -a/6, a/6]L$  or  $[a/3, a/3, -a/3]L$  and the right-hand edge bounded by a partial dislocation with the negatives of these faults. Thus there would be a dissociation

$$\frac{a}{2} [11\bar{1}] = \frac{a}{3} [11\bar{1}]L + \frac{a}{6} [11\bar{1}]R, \quad (33.23)$$

which leads to a reduction of elastic energy. A greater reduction of energy is represented by the dissociation of a screw dislocation

$$\frac{a}{2} [11\bar{1}] = \frac{a}{6} [11\bar{1}] + \frac{a}{6} [11\bar{1}] + \frac{a}{6} [11\bar{1}] \quad (33.24)$$

in which there are either three stacking faults meeting along a common line of no fault (Hirsch, 1960) or two stacking faults meeting along the central partial dislocation (Sleeswyk, 1963). These and other models formerly proposed to account for the immobility of the screw dislocation are reviewed, for example, by Hirsch (1968) and Christian and Vitek (1970), but they have been largely superseded by atomistic calculations of the core structure.

The results of various calculations all show that the screw dislocation core in a b.c.c. metal has a three-dimensional structure with threefold symmetry. The centre of the core must lie along one of the two threefold screw axes of symmetry (Suzuki, 1968), and this has the effect of either removing or reversing the spiral stacking sequence of neighbouring  $\langle 111 \rangle$  atom rows, depending on the relation between the site and the sign of the Burgers vector. The lowest-energy configurations of the stress-free dislocation are in those sites where this spiral order is reversed, and there are then two possible configurations for each dislocation which are related by a  $\langle 110 \rangle$  diad axis. The configurations have been displayed in various ways of which the most useful seems to be a differential displacement map (Vitek *et al.*, 1970) and a stress field representation (Basinski *et al.*, 1971). The differential displacement map (Fig. 7.21) shows that the largest displacements are along the three  $\{110\}$  planes intersecting the centre of the dislocation, but in each case the displacements are

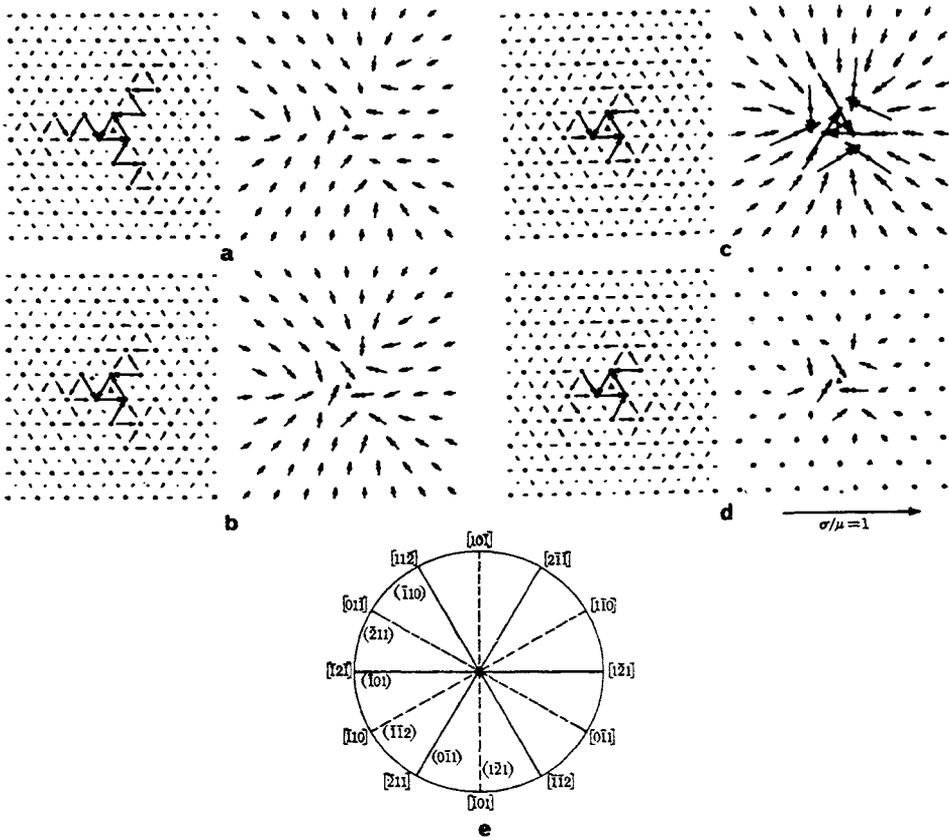


FIG. 7.21. Asymmetric core configurations computed for the unstressed  $a/2$   $[111]$  screw dislocation in the b.c.c. structure.

Displacement maps (on the left) and stress fields are shown in (a)–(c) for three different Johnson-type potentials and in (d) for the sodium potential of Fig. 5.5. The atomic structure in three successive  $(111)$  planes is projected along  $[111]$  (see (e)) and the arrows in the displacement maps are centred about the mid-points of atom pairs on successive planes and indicate the change in the  $[111]$  separations of these pairs caused by the introduction of the dislocation. The arrows are scaled so that the largest differential displacements of  $(a/6)[111]$  are represented by arrows which join two projected atom positions. The stress field map shows in the same projection the magnitude and the plane of the maximum shear stress  $\sigma$  in the  $[111]$  direction acting at each atom; the scale of this representation is given by the length of the  $\sigma/\mu = 1$  arrow.

large on only one side of the dislocation. The centre of the dislocation can be loosely regarded as the intersection of three stacking faults of generalized type (non-constant fault vector). This is also shown by the stress field map, which represents the plane and magnitude of the maximum shear stress at each atom. The stress field (Fig. 7.21) would be radial in the elastic solution, but it can be seen that with the exception of the three nearest neighbour atom rows, the arrows appear to radiate from three symmetrically placed sites away from the centre of the dislocation.

These observations on the core structure are closely connected with the asymmetry of the flow stress, which on  $\{112\}$  planes is smaller in the sense of shear which corresponds to a twinning shear than in the opposite sense, and with other anomalous slip observations in b.c.c. metals at low temperatures. The behaviour of the dislocation under stress is complex; see Basinski *et al.* (1971, 1972) and Duesbery *et al.* (1973).

Finally, we must recognize that many b.c.c. metals are elastically quite anisotropic with values of the Zener constant (Table IV. p. 76) which are especially large for the alkali metals and for  $\beta$ -brass. Head (1967) calculated, for example, that in  $\beta$ -brass dislocations with  $[111]$  Burgers vectors in a  $(1\bar{1}0)$  plane should have two instability ranges of  $\varphi$ , namely  $-26$  to  $+33^\circ$  and  $88$ – $116^\circ$  from pure screw orientation, whilst there are three instability ranges for  $[111]$  Burgers vectors in  $(11\bar{2})$  and for  $[001]$  Burgers vectors in  $(010)$ . Electron-microscopy shows that there are many V- or Z-shaped dislocations in this ordered b.c.c. structure. For less anisotropic metals like iron, the ordinary slip dislocations are stable in all orientations on  $(1\bar{1}0)$  and  $(11\bar{2})$ , but  $[001]$  dislocations have instability ranges on  $(1\bar{1}0)$  and  $(010)$ . Identification of Burgers vectors by electron-microscopy in b.c.c. metals is often difficult because of complex diffraction conditions (France and Loretto, 1968).

### 34. DISTRIBUTIONS OF DISLOCATIONS

In discussing phenomena involving point defects, we need to specify the density or atomic concentration of each defect, which may vary from point to point within the crystal. Clearly, we need a similar measure of the total amount of dislocation line present in each small volume of a crystal as a first stage in providing a full description of the dislocated state. Since a dislocation is a line defect, a specification of a scalar density alone will generally be insufficient to determine the properties of the crystal, which will depend on the types of dislocation and the ways in which they are arranged. A more complete description thus requires the specification of a tensor dislocation density at each point.

The scalar dislocation density is usually defined as the total length of all dislocation lines contained in unit volume of the crystal, and has units of  $\text{m}^{-2}$ . If the dislocations are all parallel straight lines, this density is simply the number of dislocations cutting unit area normal to the lines, and this was the definition used in the early days of dislocation theory. For dislocations which are arranged randomly, the dislocation density as defined above is about three times the number of dislocations per unit area threading through any randomly orientated planar surface in the material. When dislocations are concentrated into "walls", or nearly planar regions, it is sometimes convenient to define a line density as the number of dislocations crossing unit length of a line in the plane of the dislocations. We discuss this further in Chapter 8 which deals with crystal boundaries.

Many experimental methods are available for estimating dislocation densities, ranging from direct counts of dislocations seen in the transmission electron-microscopy of thin films, or of dislocation nucleated etch pits on the surface, to deductions from X-ray misorientations. All of these methods give reasonably self-consistent values, the densities in well-annealed pure metals being usually in the range  $10^4$ – $10^6$  lines  $\text{mm}^{-2}$ . In exceptional circumstances much lower densities may be obtained, especially in semi-conductors such as silicon

and germanium grown under carefully controlled conditions. Large crystals of germanium have been obtained virtually free of dislocations, except possibly for very small dislocation loops. Metals in "whisker" form can also be obtained substantially free from dislocations, or with a single axial screw dislocation, and macroscopic crystals of metals have been grown with densities as low as 1–10 dislocations  $\text{mm}^{-2}$  (Young, 1962; Wittels *et al.*, 1962).

A dislocation density of  $10^6$  lines  $\text{mm}^{-2}$  means that on the average about one dislocation threads through an area of about one square micron. The average dislocation separation, or the mesh size of a dislocation network, is thus of the order of 1  $\mu\text{m}$  in a well-annealed metal. It follows that dislocation effects are generally likely to be unimportant in phenomena on a scale much finer than this, since the chances are that there will be no dislocations in the region considered. This means, for example, that very fine precipitate particles, in the  $10^{-2}$ – $10^{-1}$   $\mu\text{m}$  (100–1000 Å) size range will not generally contain dislocations, and will have the properties of ideal materials.

As we have already noted, quenched metals may contain much higher defect concentrations than annealed and slowly cooled metals. Consider, for example, the tetrahedral defects formed in quenched f.c.c. metals of low stacking fault energy, which were described on p. 304. The experimental results for gold showed a typical volume density of tetrahedra of about  $5 \times 10^{21}$   $\text{mm}^{-3}$ , each tetrahedron corresponding to the condensation of about 7400 vacancies. This gives a total vacancy concentration at the quenching temperature of about  $6 \times 10^{-5}$ , in agreement with the estimates in Section 17. The equivalent dislocation density of the stair-rod dislocations in the fault tetrahedra is about  $10^8$   $\text{mm}^{-2}$ , compared with a dislocation density of  $5 \times 10^7$   $\text{mm}^{-2}$  which would have resulted if the vacancies had condensed as Frank sessile loops. In the same specimen, the density of dislocations other than those formed by vacancy condensation was estimated at  $5 \times 10^6$   $\text{mm}^{-2}$ .

Direct counts have been attempted in a few instances for the much larger dislocation densities found in heavily deformed materials, but the work is tedious and the results rather uncertain. Dislocation densities in such specimens are usually obtained from less reliable indirect methods, based on the measurement of such physical properties as the stored energy of cold work, the extra thermal or electrical resistivity produced by deformation, or the breadths of X-ray diffraction lines. The limiting densities in severely worked metals are generally found to be  $\sim 10^9$ – $10^{10}$  lines  $\text{mm}^{-2}$ , the upper limit corresponding to about one atom in a thousand being at the centre of the core of a dislocation line.

If a well-annealed single crystal attained thermodynamic equilibrium, it would contain virtually no dislocation lines. Once these lines have been introduced as inevitable accidents of growth, a mechanism for their complete removal does not exist, so a true equilibrium state is not attained, even at temperatures near to the melting point. However, if the dislocations are able to move, we expect them to adjust their positions so as to reduce their energy as much as possible. The dislocations will be in metastable equilibrium with the structure if they have adopted a configuration which gives a relative minimum to the free energy of the crystal, so that any further small displacements are resisted. This may lead to the formation of a three-dimensional dislocation network with stable nodes, as described on p. 248, or it may lead to the assembly of dislocations into stable two-dimensional networks, which then form the boundaries between slightly misaligned sub-grains. There is

experimental evidence for both these situations in different materials or different circumstances, but the commonest configuration is found to be that in which most dislocations are in two-dimensional arrays, with a smaller number of dislocations forming a three-dimensional (Frank) network distributed across the subgrains. This result was deduced first from X-ray observations, but has been amply confirmed by electron-microscopy.

When a metal is deformed, the distribution of dislocations depends upon the type of deformation, the metal concerned, and the temperature, and very many variations are possible. Experimental observations lead to the conclusion that the distributions are qualitatively similar in the majority of close-packed metals with the exception of those having very low stacking fault energy. A deformed metal does not usually contain a uniform distribution of dislocations, and those dislocations which are observed, e.g. by thin foil electron-microscopy, are often not well concentrated on to particular slip planes. In the easy glide region of single crystal deformation, the predominant structures observed are dipoles and multipole clusters of primary dislocations in mainly edge orientations. Later these clusters build up into tangles which include dislocations of several secondary systems, and eventually carpets and walls of dislocations link together in a pronounced substructure in which cells of relatively low dislocation density are separated from each other by regions of very high dislocation density. In metals deformed at sufficiently low temperatures, these dense tangles are thick dislocation walls rather than genuine sub-boundaries, but the cells on either side of a particular tangle are misorientated with respect to each other. As the temperature of deformation is raised, or if a deformed metal is heated to a higher temperature, two effects can be noticed. The first is a general cleaning of the dislocation structure in the regions of low density, and the second is a sharpening of the tangled regions until ultimately they correspond to effectively two-dimensional distributions representing sharp sub-boundaries. We shall discuss these changes further in Part II, Chapter 19; for the present, we note that the sharpness of the substructure seems to be correlated with the ease of cross-slip in the material. Sharp sub-structures are found in aluminium, which has a high stacking fault energy, allowing dislocations to cross-slip readily, even after deformation at quite low temperatures. The sub-structures in the noble metals with lower fault energies are not so sharp unless the material is annealed at a high temperature, and the scale of the sub-structure is usually too small to be detected by X-ray diffraction methods.

Alloys of very low stacking fault energy show extended dislocations lined up on slip planes in the early stages of deformation, and this is the only case in which the observations made in the ordinary way approximate to the classical "pile-ups" assumed in many theories. However, it has often been suggested that pile-ups disappear through relaxation processes on removal of stress or thinning of the foil, and in order to overcome this effect Mughrabi (1968) deformed copper crystals at 77 K and then irradiated them at 20 K without removal of stress in order to pin the dislocations in position. Subsequent electron-microscopy showed piled-up groups of dislocations containing 10–20 dislocations of the same sign.

Body-centred cubic metals develop dislocation substructures similar to those of f.c.c. metals after deformation at moderately high temperatures (e.g. room temperature for pure crystals of most of the high melting point transition metals). At low temperatures, however, very different structures are developed. The predominant features are now long screw

dislocations which are fairly uniformly distributed and tend to alternate in type, so that the long-range stress field is minimized. These screw dislocations often contain long jogs, and are accompanied by small dipoles or loop debris. This deformation substructure is believed to be associated with a much lower mobility of screws in comparison with non-screw dislocations at low temperatures.

The theory of equilibrium networks of dislocations has been developed by many workers (e.g. Frank, 1955; Ball and Hirsch, 1955; Amelinckx, 1957). We shall discuss the theory of dislocation boundaries in the next chapter, but we shall not analyse the numerous possibilities for particular types of boundary built up from square or hexagonal arrays of dislocations. Many beautiful examples of such networks in transparent crystals have been studied experimentally by precipitation techniques (Hedges and Mitchell, 1953; Amelinckx, 1957, 1958; etc.) and in foils by thin-film electron-microscopy (e.g. Carrington *et al.* 1960), and these have been correlated with theoretical predictions.

We now turn to a more mathematical description of the dislocation condition of a crystal, and we shall give a very sketchy outline of the theory of continuous distributions of dislocations, which includes the definition of a tensor dislocation density. When discussing the stress field of an individual dislocation, we made the approximate assumption that the integration can be cut off at some effective radius of the order of a mean dislocation spacing. In general, however, an arbitrary distribution of dislocations will give stresses which are not zero when averaged over distances much larger than the mean spacing, and the attainment of a configuration which reduces the free energy will correspond to a reduction in the far-reaching stresses. This leads us to consider the dislocation arrangements which are compatible with given imposed restraints, e.g. in the way in which a crystal is deformed. The first analysis of this kind was due to Nye (1953) who assumed that there is no accumulating long-range stress, i.e. that the above average is zero. Nye's work has been greatly extended by Bilby and co-workers, by Kondo, and by Kröner, who have independently and on slightly different lines developed formal theories of the compatibility relations between the overall (shape) deformation, the lattice deformation and the deformation caused by movement of dislocation lines.

We begin with a discussion of the tensor dislocation density, first introduced by Nye. Consider the dislocation lines threading through a unit area perpendicular to the  $x_j$  axis, and let the sum of the components of the Burgers vectors of these dislocation lines in the  $x_i$  direction be written  $A_{ij}$ . The quantity  $A_{ij}$  is a dislocation tensor, and it follows that the resultant Burgers vector of the dislocation lines threading a planar area  $O$  with a unit normal  $\mathbf{n}$  is given by

$$b_i = A_{ij}O_j, \quad (34.1)$$

where the vector area  $\mathbf{O}$  has components  $O_i = O n_i$ . Now let  $C$  be a closed curve forming the limit of the area  $O$ , and let  $S$  be any cap ending on  $C$ . All the dislocations threading through  $O$  must also thread through  $S$  (they may have combined or dissociated, but this is immaterial), so that we have the more general expression for the resultant Burgers vector

$$b_i = \iint_S A_{ij} dS_j \quad (34.2)$$

and  $b_i$  is independent of the choice of  $S$ .

Clearly in a real crystal if the curve  $C$  is large, we obtain the average value of the dislocation density over a region of material, whereas if it is small,  $A_{ij}$  must show large fluctuations since the dislocations are discrete. It is convenient mathematically to treat the dislocation distribution as continuous by considering an element of the crystal containing several dislocation lines, and then allowing the number of dislocation lines of each type to tend to infinity whilst the Burgers vector of each tends to zero in such a way that the product remains finite. We thus arrive at the concept of the continuously dislocated state.

When we have a continuous distribution of dislocations, we expect to be able to define a tensor dislocation density at each point, although there are no individual dislocation lines. There is now a difficulty in finding the resultant Burgers vector by using the closure failure of a circuit in the reference crystal (p. 244), since there is no good crystal in which to make the closed circuit of the real crystal, and in the strict sense there is no lattice. Bilby *et al.* (1955) have shown how this difficulty may be overcome by making use of a local correspondence between the real crystal and the reference crystal.

At any point of the real crystal, choose three independent basic vectors,  $e_i$ , which correspond to a set of basic vectors,  $a_i$ , of the reference crystal in the sense that an identification of each  $e_i$  with a corresponding  $a_i$  may be made. This choice must be made continuously at each point of the real crystal, so that  $e_i$  are always the same crystallographic vectors, and define the local crystal lattice. To an observer moving in the real crystal, the local  $e_i$  vectors are everywhere "parallel", and any two "parallel" vectors (defined by reference to the local lattice) have the same  $e_i$  components. Bilby and Smith (1956) illustrate this point graphically by the example of a number of aircraft each flying due north along a great circle of longitude. The pilots of these aircraft, using their local systems of reference, will consider that they are flying parallel to each other, but this will not be the view of an outside observer, using a Euclidean orthonormal reference system.

The local vectors  $e_i$  may be regarded as generated from the reference vectors  $a_i$  at each point of the crystal by a deformation

$$e_i = D_{ij}a_j, \tag{34.3}$$

where the components  $D_{ij}$  of the matrix  $D$  vary from point to point in the crystal. Let us now introduce a fixed orthonormal Cartesian coordinate system  $x_i$ , which is defined by a set of orthonormal vectors  $i_i$  and which we shall use to discuss displacements in the real crystal.† The set of base vectors  $i_i$  may be directly related to the lattice vectors  $a_i$  of the reference crystal; in the case of simple cubic systems, the two bases will coincide. A displacement in the real crystal from  $x_i$  to  $x_i + dx_i$  can thus be written as a vector  $dx_i i_i$ . If  $C$  is now a small closed circuit in the real crystal, we have

$$\int_C i_i dx_i = 0. \tag{34.4}$$

† The following development could be carried out in a coordinate system defined by the lattice base,  $a_i$ , or in generalised curvilinear coordinates, as shown by Bilby *et al.*, but the use of an orthonormal basis is simpler for our purpose.

Let the vectors  $\mathbf{j}_i$  be the local equivalent of the orthonormal set  $\mathbf{i}_i$ , that is let them be obtained from  $\mathbf{i}_i$  by the local deformation  $\mathbf{D}$  of eqn. (34.3). Then eqn. (34.4) becomes

$$\int_C D_{ij}^{-1} \mathbf{j}_j dx_i = 0. \quad (34.5)$$

We now require the closure failure of the corresponding circuit in the reference lattice. Each vector  $\mathbf{j}_k$  of the real lattice is replaced by its corresponding vector  $\mathbf{i}_k$  of the reference lattice, so that the closure failure giving the net Burgers vector of the distribution encircled by  $C$  is the negative of the vector sum of the reference lattice displacements, i.e.

$$\mathbf{b} = - \int_C D_{ij}^{-1} \mathbf{i}_j dx_i. \quad (34.6)$$

The sign of this equation has been written so that the Burgers vector is the vector needed to complete the circuit when this is traversed in a right-handed sense; this is obviously just a matter of defining a convention (see p. 245). In order to compare directly with eqn. (34.2), we now use Stoke's theorem to transform the line integral over  $C$  into a surface integral over any cap  $S$  having  $C$  as its limit. This gives

$$\mathbf{b} = - \iint_S \varepsilon_{ikl} \frac{\partial D_{ij}^{-1}}{\partial x_l} \mathbf{i}_j dS_k, \quad (34.7)$$

where  $\varepsilon_{ikl}$  is +1 or -1 according to whether  $i, k, l$  is an even or odd permutation of 1, 2, 3, and zero otherwise. Comparing with (34.2), we see that

$$A_{ij} = - \varepsilon_{jki} (\partial D_{li}^{-1} / \partial x_k). \quad (34.8)$$

We may think of the matrix  $\mathbf{D}$  in two equivalent ways. It establishes a one-to-one correspondence between vectors of the reference lattice and the local lattice vectors drawn from a point, so that one local region may be mapped in another. Alternatively, we may regard the lattice in an infinitesimal region of real crystal around a point at which  $D_{ij}$  is given as formed by a deformation of a corresponding region of perfect lattice. The quantities  $D_{ij}$  have thus been called both lattice correspondence functions and generating deformations. They are not, of course, correspondence matrices of the type discussed on p. 58; since the base vectors are related by the deformation  $\mathbf{D}$ , the correspondence matrix is  $\mathbf{I}$ . This only means that corresponding vectors in the real and reference crystals have the same components in coordinate systems based on  $\mathbf{e}_i$  and  $\mathbf{a}_i$  (or  $\mathbf{j}_i$  and  $\mathbf{i}_i$ ) respectively.

Small lattice vectors about a point in the real crystal may be written as  $dy_k \mathbf{j}_k$ , where  $dy_i$  is a system of local coordinates based on the vectors  $\mathbf{j}_i$ . It follows from (34.3) that the relation between the local and the reference coordinates may be written

$$dy_i = D_{ji}^{-1} dx_j, \quad dx_i = D_{ji} dy_j \quad (34.9)$$

(cf. eqns. (6.1) and (6.4)).

The Burgers vector of the circuit  $C$  could equally well have been written  $b_i = -\int_C (dy_i - dx_i)$ , so that

$$\begin{aligned} b_j &= -\int_C (D_{ij}^{-1} - \delta_{ji}) dx_i \\ &= -\int_C (\delta_{ji} - D_{ij}) dy_i. \end{aligned}$$

On using Stokes' theorem, the first expression on the right gives eqn. (34.7) again, whilst the second gives

$$b_j = -\iint_S \varepsilon_{ikl} \frac{\partial D_{ij}}{\partial y_l} dS_k, \tag{34.10}$$

which corresponds to

$$A_{ij} = -\varepsilon_{jkl} (\partial D_{li} / \partial y_k). \tag{34.11}$$

Equation (34.8) was derived first in the form (34.11) by Bilby (1955).

The relations (34.9) (known as Pfaffian forms) imply that the values of  $y_i$  at any point  $Q$  may be found from their values at any other point  $P$  by integrating along some path from  $P$  to  $Q$ . In general, however, the integral depends upon the path, so that  $y_i$  are not functions of  $x_i$ . This is an expression of the fact that we can only define a local correspondence in a dislocated crystal (see p. 244); we cannot generate the whole of the dislocated lattice from the reference lattice by applying a continuous deformation for which a displacement function exists. Imagine the reference crystal cut into small volumes, each of which is then given a local deformation specified by  $D_{ij}$ . If the separate elements can then be fitted together again so as to form a continuous lattice in ordinary space, the deformations are said to be compatible, and a continuous deformation exists. The condition for this is simply that

$$(\partial D_{ii}^{-1} / \partial x_k) = (\partial D_{ki}^{-1} / \partial x_i), \tag{34.12}$$

and we see from (34.7) or (34.10) that this corresponds to zero dislocation density.

The condition for compatibility, or for a continuous mapping of the reference lattice in the dislocated crystal, is thus the absence of dislocations. Conversely, when the local deformations of the separate elements imagined above are not compatible, dislocations will be needed to fit these elements together in real space.

Equation (34.7) gives the Burgers vector associated with the circuit element terminating the surface  $S$ , and should be distinguished from the *local Burgers vector*, which is the vector of the real lattice corresponding to the vector  $\mathbf{b}$  of the reference crystal. The local Burgers vector could be obtained, for example, by completing a circuit in the reference crystal and taking the negative of the closure failure of the corresponding circuit in the real crystal (see p. 245). By definition, the local Burgers vector will have the same components in the  $\mathbf{j}_k$  system as the Burgers vector has in the  $\mathbf{i}_k$  system, so that it may be written

$$\begin{aligned} \mathbf{l} &= \iint_S A_{ij} D_{ik} \mathbf{j}_k dS_j \\ &= -\iint_S \varepsilon_{jkl} D_{im} \frac{\partial D_{ii}^{-1}}{\partial x_k} \mathbf{i}_m dS_j. \end{aligned} \tag{34.13}$$

The geometrical significance of the local Burgers vector is that if a dislocation line moves in the real crystal by glide or climb so as to intersect the closed circuit  $C$ , the parts of the circuit on each side of the point of intersection suffer a relative displacement of  $d\mathbf{l}$ , where  $d\mathbf{l}$  is the change in the local Burgers vector at the point of intersection. Clearly the distinction between  $\mathbf{b}$  and  $\mathbf{l}$  is unimportant when  $\mathbf{D}$  represents a small deformation; it becomes important when there are large discontinuities in the tensor dislocation density, as in the theory of surface dislocations (p. 365).

Although we have written the area element as a vector in the above equations, it is strictly an axial vector, that is an asymmetric second order tensor. The tensor element of area is related to the (pseudo) vector element of area by the relation

$$\left. \begin{aligned} dS_k &= \frac{1}{2} \varepsilon_{kmn} dS_{mn}, \\ dS_{mn} &= \varepsilon_{kmn} dS_k. \end{aligned} \right\} \quad (34.14)$$

On substituting into (34.13), we find that the local Burgers vector can be written in the form

$$l_p = T_{mnp} dS_{mn}, \quad (34.15)$$

where

$$T_{mnp} = \frac{1}{2} D_{jp} \left\{ \frac{\partial D_{mj}^{-1}}{\partial x_n} - \frac{\partial D_{nj}^{-1}}{\partial x_m} \right\}. \quad (34.16)$$

Alternatively, we may define a local dislocation tensor density  ${}^l A_{ij}$  where

$$l_j = {}^l A_{ij} dS_j \quad (34.17)$$

and

$${}^l A_{ij} = \varepsilon_{jkl} T_{ikl}. \quad (34.18)$$

The quantity  $T_{ijk}$  is called the torsion tensor.

In the situation considered by Nye, in which there are no far reaching stresses, the relation between  $\mathbf{i}_i$  and  $\mathbf{j}_i$  is everywhere a pure rotation, so that the real lattice although continuously curved is unstrained. The general theory has been applied to this case, and to many other specific examples (for review see Bilby, 1960), but we shall not consider them in detail. The further development of the theory makes extensive use of the geometry of generalized spaces in which the lattice generating deformations are compatible. We recall that the individual elements of a reference crystal, given arbitrary deformations, will not fit together without the introduction of further deformations or the insertion of dislocations between the elements. However, the elements may be "fitted together" in a general space with appropriate geometry, so that the continuously dislocated crystal may be associated with such a space. The advantage of this formalism is that the previously developed methods of differential geometry for such a space may be used to solve dislocation problems.

The analogy with the theory of generalized spaces may be introduced by considering equivalent lattice vectors at two neighbouring points  $P$  and  $Q$ . Clearly, two vectors are crystallographically equivalent if they have the same components relative to the local basis ( $\mathbf{e}_i$  or  $\mathbf{j}_i$ ). Let two such vectors at  $P$  and  $Q$  be written  $p_i \mathbf{i}_i = p_k D_{ki}^{-1}(P) \mathbf{j}_i$  and  $(p_i + dp_i) \mathbf{i}_i =$

$(p_k + dp_k) D_{ki}^{-1}(Q) \mathbf{j}_i$ , where we have found it necessary to distinguish the values of  $D_{ki}^{-1}$  at  $P$  and  $Q$ . Since the components are identical in the  $\mathbf{j}_i$  system, we have

$$p_k D_{ki}^{-1}(P) = (p_k + dp_k) D_{ki}^{-1}(Q)$$

or, since  $P$  and  $Q$  are neighbouring points,

$$dp_k D_{ki}^{-1} = -p_k \frac{\partial D_{ki}^{-1}}{\partial x_l} dx_l.$$

This may be written

$$dp_m = -D_{im} \frac{\partial D_{ki}^{-1}}{\partial x_l} p_k dx_l = -L_{klm} p_k dx_l. \quad (34.19)$$

In differential geometry, a relation like (34.19) is described as a linear connection; it prescribes which vectors of the generalized space are equivalent or "parallel". The quantities  $L_{klm}$  are called the coefficients of connection, and we see that they may be written

$$L_{klm} = D_{im} \frac{\partial D_{ki}^{-1}}{\partial x_l}. \quad (34.20)$$

The important geometrical property of this connection is that  $L_{klm}$  is antisymmetric in  $k$  and  $l$ , i.e.  $L_{klm} \neq L_{lkm}$ . For a Euclidean geometry,  $L_{klm} = 0$ , and in a more general (Riemannian) space,  $L_{klm}$  is still symmetric. A space for which  $L_{klm}$  is not symmetric is said to possess torsion, and its geometry is described as non-Riemannian. The quantities  $L_{klm}$  may be written as the sum of a symmetric and an antisymmetric part; by comparing with eqn. (34.16), we see that the antisymmetric part gives the torsion tensor associated with the local Burgers vector since

$$T_{klm} = \frac{1}{2}(L_{klm} - L_{lkm}). \quad (34.21)$$

The unusual geometrical properties of a space with torsion are just those properties which we associate with a crystal containing dislocations if we imagine ourselves inside the crystal with only the lattice lines to guide us. For example, an infinitesimal closed parallelogram does not exist in a space with torsion. If we draw two non-parallel small vectors  $PQ$  and  $PR$  from  $P$ , and then draw  $QS$  equal to and parallel with  $PR$  and  $RS'$  equal to and parallel with  $PQ$ , in general  $S$  does not coincide with  $S'$ . This is just the situation in a dislocated crystal, the vector  $SS'$  being, of course, the local Burgers vector encircled by the circuit.

We conclude this section by referring again to the problem of the generation of a dislocated crystal from a reference crystal. We have emphasized that the local lattice generating deformations are not compatible in the sense that they will not fit together in ordinary space without the introduction of discontinuities (dislocations). We find it convenient to distinguish between these deformations and the *shape deformation*, which could be measured in principle by the distortion of a network of lines scribed on the crystal before the introduction of the dislocations. The shape deformation describes the behaviour of large vectors, and is the deformation usually considered in the macroscopic theory of plasticity. Since the final crystal may be regarded as obtained by the deformation of a reference crystal in ordi-

nary space, the shape deformation is necessarily compatible. This means that the deformation is continuous and the displacements are the derivatives of a displacement function.

In most of this section we have been concerned with the geometry of an existing dislocated crystal, and for this reason we have not considered the shape deformation. In such a crystal the lattice strain is clearly related to the dislocation density, as we have seen. However, when we consider the production of the dislocated crystal from the reference crystal, it is obvious that there may be changes of shape which do not affect the lattice at all. Such a change, for example, is produced by the generation of a dislocation line which passes right through the crystal and disappears again at the surface. In general, all dislocation motion by glide or climb will produce changes in the external shape of the body.

We may thus regard the shape deformation as made up of two parts, one due to changes in the lattice (the *lattice deformation*), and one due to the introduction and movement of dislocations without affecting the lattice. This second part has been called the *lattice invariant deformation* (Bilby and Christian, 1956) or the *dislocation deformation* (Bilby and Smith, 1956). In terms of our previous discussion, we apply separate lattice deformations to individual elements of the reference crystal, and we now make these elements fit together in real space. We can do this by causing slip to occur in the elements and by adding or removing lattice planes to the elements. Since these operations will differ from element to element, lattice dislocations will appear when the elements are fitted together. The operations required to make the elements fit together in ordinary space constitute the lattice invariant deformation. We pass from the discrete (physically real) dislocation model to the continuous (mathematically convenient) model by making the elements infinitesimal. The torsion calculated from  $D_{ij}$  then gives just the dislocation distribution required to make the resultant shape changes of the elements compatible with each other.

We have regarded the shape change as continuous in the sense that two points which are initially very close together remain very close together. However, some of the more important applications of the above theory correspond to surfaces of discontinuity in the dislocation field, where the shape and lattice deformations suddenly change. Such a change represents the boundary between two crystals of different structure and orientation, and when a lattice correspondence exists, the dislocation distribution at the interface ensures compatibility of the two shape deformations on each side of the interface. The dislocation distribution<sup>†</sup> thus ensures that macroscopic regions of the two crystals will fit smoothly together, even although the lattices will not. We discuss the application of the theory of continuous distributions of dislocations to two-dimensional sheets ("surface dislocations") in Section 38. A fuller description of the relation between lattice and shape deformations in the theory of martensite is given in Part II, Chapters 21 and 22.

#### REFERENCES

- AMELINCKX, S. (1957) *Dislocations and Mechanical Properties of Crystals*, p. 2, Wiley New York; (1958) *Acta metall.* **6**, 34.  
 AMELINCKX, S., BONTINCK, W., DEKEYSER, W., and SEITZ, F. (1957) *Phil. Mag.* **2**, 355.  
 BABUSKA, I., VITÁSEK, E., and KROUPA, F. (1960) *Czech. J. Phys. B*, **10**, 419, 488.

<sup>†</sup> Or the equivalent effect produced by alternating fine twins: see Part II, Chapter 22.

- BACON, D. J. (1967) *Phys. Stat. Sol.* **23**, 527; (1969) *Scripta metall.* **3**, 735.
- BACON, D. J., BARNETT, D. M. and SCATTERGOOD, R. O. (1978) *Progr. in Mater. Sci.* **23**, 51.
- BACON, D. J., BULLOUGH, R., and WILLIS, J. R. (1970) *Phil. Mag.* **22**, 31.
- BALL, C. J. and HIRSCH, P. B. (1955) *Phil. Mag.* **46**, 1343.
- BARDEEN, J. and HERRING, C. (1952) *Imperfections in nearly Perfect Crystals*, p. 261, Wiley, New York.
- BARNETT, D. M. (1972) *Phys. stat. sol.* (b), **49**, 741.
- BARNETT, D. M., and SWANGER, L. A. (1971) *Phys. stat. sol.* **48**, 419.
- BARNETT, D. M., ASARO, R. J., BACON, D. J., GAVAZZA, S. D., and SCATTERGOOD, R. O. (1972) *J. Phys.* (F) **2**, 854.
- BASINSKI, Z. S. (1959) *Phil. Mag.* **4**, 393; (1964) *Disc. Faraday Soc.* **38**, 93.
- BASINSKI, Z. S. and CHRISTIAN, J. W. (1954) *Acta metall.* **2**, 101; (1960) *Aust. J. Phys.* **13**, 299.
- BASINSKI, Z. S., DUESBERY, M. S., and TAYLOR, R. (1970) *Phil. Mag.* **21**, 1201; (1971) *Can. J. Phys.* **49**, 2160; (1972) *Interatomic Potentials and Simulation of Lattice Defects*, p. 537, Plenum Press, New York.
- BILBY, B. A. (1950) *Proc. Phys. Soc. A*, **63**, 191; (1951) *Research* **4**, 389; (1953) *Phil. Mag.* **44**, 782; (1955) *Report of the Conference on Defects in Crystalline Solids*, p. 123, Physical Society, London; (1960) *Prog. Solid Mech.* **1**, 329.
- BILBY, B. A. and CHRISTIAN, J. W. (1956) *The Mechanisms of Phase Transformations in Metals*, p. 121, Institute of Metals, London.
- BILBY, B. A. and SMITH, E. (1956) *Proc. R. Soc. A*, **236**, 481.
- BILBY, B. A., BULLOUGH, R., and SMITH, E. (1955) *Proc. R. Soc. A*, **231**, 263.
- BRAILS福德, A. D. (1965) *Phys. Rev.* **139**, A1813.
- BROWN, L. M. (1964) *Phil. Mag.* **10**, 441; (1967) *Ibid.* **15**, 363.
- BULLOUGH, R. and FOREMAN, A. J. E. (1964) *Phil. Mag.* **9**, 315.
- BULLOUGH, R. and NEWMAN, R. C. (1960) *Phil. Mag.* **5**, 921.
- BULLOUGH, R. and PERRIN, R. C. (1968) *Proc. R. Soc. A*, **305**, 541.
- BURGERS, J. M. (1939) *Proc. K. ned. Acad. Wet.* **42**, 293, 378.
- CAHN, J. W. (1960) *Acta metall.* **8**, 554.
- CARRINGTON, W., HALE, K. F., and MACLEAN, D. (1960) *Proc. R. Soc. A*, **259**, 203.
- CELLI, V. (1961) *J. Phys. Chem. Solids* **19**, 100.
- CHANG, R. (1962) *Acta metall.* **10**, 951; (1967) *Phil. Mag.* **16**, 1021.
- CHANG, R., and GRAHAM, L. J., (1966) *Phys. stat. sol.*, **18**, 99.
- CHOU, Y. T. (1965) *Acta metall.* **13**, 251, 779.
- CHRISTIAN, J. W. (1968) *The Interaction between Dislocations and Point Defects*, p. 604, HMSO, London.; (1970) *Proceedings of the 2nd International Conference on Strength of Metals and Alloys*, p. 31, American Society for Metals, Ohio; (1973) *Int. Met. Rev.* **18**, 24.
- CHRISTIAN, J. W. and SWANN, P. R. (1965) *Alloying Behaviour and Effects of Concentrated Solid Solutions*, Gordon and Breach, New York. p. 105.
- CHRISTIAN, J. W. and VÍTEK, V. (1970) *Rep. Progr. Phys.* **33**, 307.
- CLAREBROUGH, L. M. and HEAD, A. K. (1969) *Phys. stat. sol.* **33**, 431.
- CLAREBROUGH, L. M., SEGALL, R. L., and LORETTO, M. H. (1966) *Phil. Mag.* **13**, 1285.
- COCHARDT, A. W., SCHOECK, G., and WIEDERSICH, H. (1955) *Acta metall.* **3**, 533.
- COCKAYNE, D. J. H. (1972) *Z. Naturforsch.* **27a**, 452.
- COCKAYNE, D. J. H., JENKINS, M. L., and RAY, I. L. F. (1971) *Phil. Mag.* **24**, 1383.
- COCKAYNE, D. J. H., RAY, I. L. F., and WHELAN, M. J. (1969) *Phil. Mag.* **20**, 1265.
- CONRAD, H. (1963) *The Relation between the Structure and Mechanical Properties of Metals*, p. 474, HMSO, London.
- CONRAD, H. and SCHOECK, G. (1960) *Acta metall.* **8**, 791.
- COTTERILL, R. M. J. and DOYAMA, M. (1966) *Phys. Rev.* **145**, 465.
- COTTRELL, A. H. (1948) *Conference on Strength of Solids*, p. 30, The Physical Society, London; (1952) *Phil. Mag.* **42**, 1327; (1953) *Dislocations and Plastic Flow in Crystals*, Clarendon Press, Oxford; (1958) *Trans. Metall. Soc. Am. Inst. Min. (Metall.) Engrs.* **212**, 192.
- COTTRELL, A. H. and BILBY, B. A. (1951) *Phil. Mag.* **42**, 573.
- CRUSSARD, C. (1950) *Métaux Corros.* **25**, 203.
- DASH, W. C. (1956) *Dislocations and Mechanical Properties of Crystals* p. 57, Wiley, New York.
- DE JONG, M. and KOEHLER, J. S. (1963) *Phys. Rev.* **129**, 49.
- DE WIT, R. (1960) *Solid State Physics.* **10**, 249; (1967) *Phys. stat. sol.* **20**, 575.
- DE WIT, R. and KOEHLER, J. S. (1959) *Phys. Rev.*, **116**, 1113.
- DIETZE, H. D. (1952) *Z. Physik* **132**, 107.

- DOBSON, P. S. and SMALLMAN, R. E. (1966) *Proc. R. Soc. A*, **293**, 423.
- DOBSON, P. S., GOODREW, P. J., and SMALLMAN, R. E. (1967) *Phil. Mag.* **16**, 9.
- DOYAMA, M. S. and COTTERILL, R. M. J. (1966) *Phys. Rev.* **150**, 448.
- DUESBERY, M. S., VÍTEK, V., and BOWEN, D. K. (1973) *Proc. R. Soc. A*, **332**, 85.
- EDINGTON, J. W. and SMALLMAN, R. E. (1965) *Phil. Mag.* **11**, 1109.
- EDINGTON, J. W. and WEST, D. R. (1966) *Phil. Mag.* **14**, 603; (1967) *Ibid.* **15**, 229.
- ESCAIG, B. (1963) *Acta metall.* **11**, 595.
- ESHELBY, J. D. (1949) *Proc. Phys. Soc. A*, **62**, 307; (1954) *J. Appl. Phys.* **25**, 255; (1956a) *Solid State Phys.* **3**, 79; (1956b) *Proc. Phys. Soc. B*, **69**, 1013.
- ESHELBY, J. D., READ, W. T., and SHOCKLEY, W. (1953) *Acta metall.* **1**, 251.
- EVANS, A. G. and RAWLINGS, R. D. (1969) *Phys. stat. sol.* **34**, 9.
- FOREMAN, A. J. E. (1955) *Acta metall.* **3**, 322; (1967) *Phil. Mag.* **15**, 1011.
- FOREMAN, A. J. E., JASWON, M. A., and WOOD, J. K. (1951) *Proc. Phys. Soc. A*, **64**, 156.
- FRANCE, L. K. and LORETTO, M. H. (1968) *Proc. R. Soc. A*, **307**, 83.
- FRANK, F. C. (1949) *Proc. Phys. Soc. A*, **62**, 202; (1951) *Phil. Mag.* **42**, 809; (1955) *Conference on Defects in Crystalline Solids*, p. 159, The Physical Society, London.
- FRANK, F. C. and NICHOLAS, J. F. (1953) *Phil. Mag.* **44**, 1213.
- FRANK, F. C. and READ, W. T. (1950) *Phys. Rev.* **79**, 722.
- FRANK, F. C. and VAN DER MERWE, J. H. (1949) *Proc. R. Soc. A*, **198**, 205.
- FRIEDEL, J. (1964) *Dislocations*, Pergamon Press, Oxford.
- GALLAGHER P. C. J. (1966) *Phys. stat. sol.* **16**, 695; (1968) *J. Appl. Phys.* **39**, 160.
- GALLAGHER, P. C. J. and LIU, Y. C. (1969) *Acta metall.* **17**, 127.
- GEHLEN, P. C. (1970) *J. Appl. Phys.* **41**, 5165; (1972) *Interatomic Potentials and Simulation of Lattice Defects*, p. 475, Plenum Press, New York.
- GILMAN, J. J. (1960) *Aust. J. Phys.* **13**, 327.
- GROVES, P. P. and BACON, D. J. (1970) *Phil. Mag.* **22**, 83.
- HAASEN, P. (1957) *Acta metall.* **5**, 598.
- HALES, R., SMALLMAN, R. E., and DOBSON, P. S. (1968) *Proc. R. Soc. A*, **307**, 71.
- HARRIS, J. E. and MASTERS, B. C. (1966a) *Phil. Mag.* **13**, 963; (1966b) *Proc. R. Soc. A*, **292**, 240.
- HEAD, A. K. (1953) *Phil. Mag.* **44**, 92; (1967) *Phys. stat. sol.* **19**, 185.
- HEDGES, J. M. and MITCHELL, J. W. (1953) *Phil. Mag.* **44**, 223, 357.
- HEIDENREICH, R. D. and SHOCKLEY, W. (1948) *Report of the Conference on Strength of Solids*, p. 57, The Physical Society, London.
- HIRSCH, P. B. (1959a) *Conference on Internal Stresses and Fatigue*, p. 139, Elsevier, Amsterdam; (1959b) *Metall. Rev.* **4**, 113; (1960) *Fifth International Conference in Crystallography, Cambridge* (oral discussion); (1962) *Phil. Mag.* **7**, 67; (1968) *Trans. Jap. Inst. Metals*, Suppl. **9**, XXX.
- HIRSCH, P. B. and WARRINGTON, D. H. (1961) *Phil. Mag.* **6**, 735.
- HIRTH, J. P. and COHEN, M. (1969a) *Scripta metall.* **3**, 107; (1969b) *Ibid.* **3**, 311.
- HIRTH, J. P. and LOTHE, J. (1968) *Theory of Dislocations*, McGraw-Hill, New York.
- HIRTH, J. P. and NIX, W. D. (1969) *Phys. stat. sol.* **35**, 177.
- HULL, D. (1958) *Phil. Mag.* **3**, 1468; (1960) *Acta metall.* **8**, 11.
- HUMBLE, P. and FOREWOOD, C. T. (1968a) *Phys. stat. sol.* **29**, 99; (1968b) *Aust. J. Phys.* **21**, 941.
- HUNTINGTON, H. B. (1955) *Proc. Phys. Soc. B*, **68**, 1043.
- HUNTINGTON, H. B., DICKEY, J. E., and THOMSON, R. (1955) *Phys. Rev.* **100**, 1117; (1959) *Ibid.* **113**, 1696.
- INDENBOM, V. L. and ORLOV, S. S. (1967) *Kristallografiya* **12**, 971 (*Soviet Phys. Crystallogr.* **12**, 849).
- JOHNSTON, W. G. and GILMAN, J. J. (1959) *J. Appl. Phys.* **30**, 129.
- JØSSANG, T. (1968) *Phys. stat. sol.* **27**, 579.
- JØSSANG, T., LOTHE, J., and SKYLSTAD, K. (1965) *Acta metall.* **13**, 271.
- JØSSANG, T., STOWELL, M. J., HIRTH, J. P., and LOTHE, J. (1965) *Acta metall.* **13**, 279.
- KLÉMAN, M. and FRIEDEL, J. (1969) *J. Phys. Paris* **30**, C4-43.
- KOEHLER, J. S. (1941) *Phys. Rev.* **60**, 397.
- KRONBERG, M. L. (1959) *J. Nucl. Mater.* **1**, 85; (1961) *Acta metall.* **9**, 970.
- KRÖNER, E. (1953) *Z. Phys.* **136**, 402; (1958) *Kontinuums theorie der Versetzungen und Eigenspannungen*, *Ergebn. Angew. Math.* **5**, 1.
- KROUPA, F. (1960) *Czech. J. Phys. B*, **10**, 284; (1966) *Theory of Crystal Defects*, p. 275, Academia, Prague.
- KUHLMANN-WILSDORF, D. (1958) *Phil. Mag.* **3**, 125; (1960) *Phys. Rev.* **120**, 773; (1962) *Strengthening Mecha-*

- nisms in Solids, p. 137, American Society for Metals, Cleveland, Ohio; (1962) *Trans. metall. Soc. AIME* **224**, 1047; (1965) *Acta metall.* **13**, 257.
- LALLY, J. S. and PARTRIDGE, P. G. (1966) *Phil. Mag.* **13**, 9.
- LEIBFRIED, G. (1950) *Z. Phys.* **127**, 344.
- LEWTHWAITE, G. W. (1966) *Phil. Mag.* **13**, 437.
- LOMER, W. M. (1951) *Phil. Mag.* **42**, 1327; (1957) *Ibid.* **2**, 1053.
- LORETTO, M. H., CLAREBROUGH, L. M., and SEGALL, R. L. (1965) *Phil. Mag.* **11**, 459.
- LOTHE, J. (1967) *Phil. Mag.* **15**, 353.
- LOUAT, N., (1956) *Proc. Phys. Soc. B*, **69**, 459.
- LOW, J. R. and TURKALO, A. M. (1962) *Acta metall.* **3**, 215.
- MALEN, K. (1970) *Phys. stat. sol.* **38**, 259.
- MALEN, K. and LOTHE, J. (1970) *Phys. stat. sol.* **39**, 287.
- MARADUDIN, A. A. (1958) *J. Phys. Chem. Solids* **9**, 1.
- MASTERS, B. C. and CHRISTIAN, J. W. (1964) *Proc. R. Soc. A*, **281**, 223.
- MILTAT, J. E. A. and BOWEN, D. K. (1970) *Phys. stat. sol. (a)* **3**, 431.
- MORTON, A. J. and CLAREBROUGH, L. M. (1969) *Aust. J. Phys.* **22**, 393.
- MOTT, N. F. and NABARRO, F. R. N. (1948) *Report of the Conference on Strength of Solids*, p. 38, The Physical Society, London.
- MUGHRABI, H. (1968) *Phil. Mag.* **18**, 1211.
- NABARRO, F.R.N. (1947) *Proc. Phys. Soc.* **59**, 256; (1951) *Proc. R. Soc. A*, **209**, 278; (1952) *Adv. Phys.* **1**, 319; (1967) *Theory of Crystal Dislocations*, Clarendon Press, Oxford.
- NABARRO, F. R. N., BASINSKI, Z. S. and HOLT, D. B. (1964) *Adv. in Phys* **13**, 193.
- NYE, J. F. (1953) *Acta metall.* **1**, 153.
- PEACH, M. O. (1951) *J. Appl. Phys.* **22**, 1359.
- PEACH, M. O. and KOEHLER, J. S. (1950) *Phys. Rev.* **80**, 436.
- PEIERLS, R., (1940) *Proc. Phys. Soc.* **52**, 34.
- READ, W. T. and SHOCKLEY, W. (1952) *Imperfections in Nearly Perfect Crystals*, p. 77, Wiley, New York.
- SAADA, G. (1962) *Acta metall.* **10**, 551.
- SAADA, G. and WASHBURN, J. (1963) *J. Phys. Soc. Japan* **18**, Suppl. **3**, 43.
- SAXL, I. (1964) *Czech. J. Phys. B*, **14**, 381; (1968) *Czech. J. Phys. B*, **18**, 39.
- SCHAPINK, F. W. and DE JONG, M. (1964) *Acta metall.* **12**, 756.
- SCHOECK, G. (1969) *Scripta metall.* **3**, 239.
- SCHWARTZKOPFF, K. (1969) *Acta metall.* **17**, 345.
- SEEGER, A. (1955a) *Phil. Mag.* **46**, 1194; (1955b) *Conference on Defects in Crystalline Solids*, p. 328, The Physical Society, London; (1964) *Discuss. Faraday Soc.* **38**, 82.
- SEEGER, A., BERNER, R., and WOLF, H. (1959) *Z. Phys.* **155**, 247.
- SEEGER, A., DONTH, H., and PFAFF, F. (1957) *Discuss. Faraday Soc.* **23**, 19.
- SEEGER, A. and SCHOECK, G. (1953) *Acta metall.* **1**, 519.
- SEIDMAN, D. N. and BALLUFFI, R. W. (1966) *Phys. stat. sol.* **17**, 531.
- SEITZ, F. (1952) *Adv. Phys.* **1**, 43.
- SHOCKLEY, W. (1947) Quoted by Mott and Nabarro (1948).
- SILCOX, J. and HIRSCH, P. B. (1959) *Phil. Mag.* **4**, 72.
- SLEESWYK, A. W. (1963) *Phil. Mag.* **8**, 1467.
- STEEDS, J. W. (1967) *Phil. Mag.* **16**, 771, 785.
- STEHLE, H. and SEEGER, A. (1956) *Z. Phys.* **146**, 217.
- STEIN, D. F. and LOW, J. R. (1960) *J. Appl. Phys.* **31**, 362; (1966) *Acta metall.* **14**, 1183.
- STOBBS, W. M. and SWORN, C. H. (1971) *Phil. Mag.* **24**, 1365.
- STROH, A. N. (1956) *Phil. Mag.* **1**, 489; (1958) *Ibid.* **3**, 625.
- SUZUKI, H. (1952) *Sci. Rep. Res. Inst. Tohoku Univ.* **A4**, 455; (1968) *Dislocation Dynamics*, p. 679, McGraw-Hill, New York.
- TAYLOR, G. I. (1934) *Proc. R. Soc. A*, **145**, 362.
- THOMPSON, N. (1953) *Proc. Phys. Soc. B*, **66**, 481; (1955) *Report of the Conference on Defects in Crystalline Solids*, p. 153, The Physical Society, London.
- TOMSETT, D. I. and BEVIS, M. (1969) *Phil. Mag.* **19**, 129.
- VÍTEK, V. (1974) *Crystal Lattice Defects* **5**, 1.
- VÍTEK, V., PERRIN, R. C., and BOWEN, D. K. (1970) *Phil. Mag.* **21**, 1049.

- WEERTMAN, J. (1965) *Phil. Mag.* **11**, 1217; (1967) *J. Appl. Phys.* **38**, 5293.  
WESTLAKE, D. G. (1961) *Acta metall.* **9**, 327.  
WESTMACOTT, K. H., BARNES, R. S., HULL, D., and SMALLMAN, R. E. (1961) *Phil. Mag.* **6**, 929.  
WHELAN, M. J. (1959) *Proc. R. Soc. A*, **249**, 114.  
WILKENS, M. (1967) *Acta metall.* **15**, 1417; (1968) *Scripta metall.* **2**, 299.  
WILLIS, J. R. (1970) *Phil. Mag.* **21**, 931.  
WITTELS, M. C., SHERRILL, F. A., and YOUNG, F. W., (1962) *Appl. Phys. Letters* **1**, 22.  
YAMAGUCHI, M. and VÍTEK, V. (1973) *F. Phys. (F)* **3**, 523, 537.  
YOUNG, F. W. (1962) *Bull. Am. Phys. Soc.* **7**, 215.  
YOFFE, E. H. (1960) *Phil. Mag.* **5**, 161.

## CHAPTER 8

# *Polycrystalline Aggregates*

### 35. MACROSCOPIC THEORY

The description of the solid state in Chapters 5 and 6 refers only to single crystals of a metal or one-phase alloy. Very large single crystals can often be grown by a suitable technique, but macroscopic specimens usually consist of a compact polycrystalline mass. The crystals are allotriomorphic, i.e. their limiting surfaces are not regular and do not display the symmetry of the internal structure. In a single-phase assembly, neighbouring crystals differ only in the orientations of their respective lattices: it is then usual to refer to the individual crystals as the grains of the structure and the regions over which the lattice orientation changes are the grain boundaries. We shall restrict the use of the term grain boundary to this sense, and regions of discontinuity separating crystals having different structures or chemical compositions will be referred to as crystal boundaries or interphase boundaries.

A one-phase assembly is in true thermodynamical equilibrium only when it forms a single crystal, the exterior surface of which has the shape giving the minimum energy. In practice the energies of the free surface and the grain boundaries of a polycrystalline specimen are very small, and there is a negligible rate of approach to this equilibrium except at temperatures near the melting point. Even at these temperatures, the structure will usually remain polycrystalline indefinitely, unless there has been prior mechanical deformation. When calculating the equilibrium state of a solid metal or alloy, we have, therefore, to accept the various crystals as frozen into the structure.<sup>†</sup> In exactly the same way, we have to accept the presence of isolated dislocations in the structure, even though they can never be in thermal equilibrium with the lattice.

In a one-phase assembly, the orientations of the grains may be completely random, or there may be a preferred orientation induced by mechanical deformation or thermal treatment. In a drawn wire, for example, there is a tendency for some particular crystallographic axis to lie along the wire axis, the various grains having random rotations about this direction. This is described as a fibre structure. The grains in a cold-rolled sheet have a preferred crystal direction in the direction of rolling, and a crystal plane in the plane of rolling. The

<sup>†</sup> Speculations have sometimes been made, e.g. Born (1946), that there may be a natural or equilibrium limit to the size of a region of perfect lattice. It seems most improbable that this can be true.

preferred orientations are never perfect, and are represented statistically by "pole figures" plotted in stereographic projection.

The thickness of the transition region between two lattices is always small, and grain boundaries may be treated macroscopically as surfaces of discontinuity. A general grain boundary has five degrees of freedom, of which three degrees are associated with the relative orientations of the two lattices on either side of the boundary, and two are required to specify the local inclination of the boundary surface relative to one of these lattices. The energy of the boundary will be a function of all these variables, but the relative orientation of the two lattices proves to be much more important than that of the boundary itself, except in certain special cases. The atoms in a grain boundary region are acted upon by forces tending to move them into positions corresponding to the two competing orientations. In general, this implies a considerable distortion of the periodic structure, and part or all of the grain boundary must be bad crystal. As we have seen, the atoms in a free surface are distorted little from equilibrium positions; nevertheless, the energy of a grain boundary is less than that of a free surface, since there are fewer unsatisfied atomic bonds in the former.

It is sometimes suggested, e.g. Chalmers and Gleiter (1971), that in addition to the above five parameters, three further quantities are required to specify the components of any relative translation  $\mathbf{t}$  of the atoms at large distances from the boundary in the two structures (cf. eqn. (9.1)), thus making eight parameters, and that a ninth may be necessary to specify the position of the boundary itself. However, this appears to confuse internal parameters which may be utilized in an economic description of the boundary structure with macroscopic degrees of freedom: only five parameters may be varied independently of the others. As discussed later in this chapter, the translation  $\mathbf{t}$  must be a vector from a lattice point of one structure to an internal point of a unit cell which has that lattice point at a corner and, together with any localized atomic adjustments or relaxations, the eight parameters fix the positions of all the atoms on both sides of the interface. There is thus no necessity separately to specify the position of the interface. Once the five degrees of freedom are fixed, there are in principle  $3N$  additional but dependent parameters to specify the positions of the  $N$  atoms; these may be reduced approximately to three parameters in many models, but they are still not independent.

The free energy of an internal or external surface is a useful macroscopic concept, which we have hitherto used without detailed justification. Such a free energy may be defined quite generally as the change in the free energy of the whole assembly per unit area of boundary formed when a geometrical boundary is introduced by some virtual process. It is important to realize that this energy is a property of the whole solid, and is not necessarily localized in the immediate vicinity of the boundary region. When the boundary is plane, the free energy is independent of the exact location of the supposed geometrical boundary within the transition region but, in common with other macroscopic surface parameters, the free energy per unit area of a curved boundary depends slightly on the way in which the boundary is defined. The important result, however, is that the thermodynamic relations between these parameters are independent of the position of the geometrical surface; a detailed discussion has been given by Herring (1953). For specimens having dimensions large compared with the effective thickness of the transition region, the variation of specific surface

energy with curvature may be neglected, and the extra energy associated with the presence of edges or corners may similarly be ignored.

Another useful macroscopic concept is that of surface stress. This may be defined by considering the difference in the forces acting across a small area normal to the boundary before and after introduction of the boundary. The surface stress in a liquid is a uniform tensile force acting normally to any line in the liquid surface, and numerically equal to the surface free energy of the liquid. This surface tension is just another measure of the tendency of the liquid to attain a minimum area surface, and may be introduced by considering the work done in increasing the area by a small amount. In general, however, the surface stress has three independent components, since solids can withstand elastic shearing forces, and the numerical equivalence of the surface free energy and the surface stress disappears. Indeed, the surface free energy is always positive, but the non-shear component of the surface stress may have either sign. This complication has led to some confusion in the terminology, and particularly in the use of "surface tension". Shuttleworth (1950) defined the surface tension as one half of the sum of the principal surface stresses, by analogy with the three-dimensional definition of hydrostatic pressure. Herring and other workers have continued to use the term surface tension to mean the surface free energy defined for a particular choice of the geometrical interface which gives exactly the classical relation between excess pressure and curvature. To avoid confusion, we shall not refer to surface tension, except for liquids.

The distinction between the surface free energy and the surface stress is that the former measures the work required to create new boundary, while the latter is a measure of the work required to increase the boundary area by deformation, without changing the number of atoms in the boundary. Brooks (1952) has shown that for internal boundaries, there are two surface stress tensors, one for each side of the boundary. Even with an external surface, there are complications arising from situations intermediate between those just envisaged. Thus, in general, the application of a stress will result in changes both in the surface area per atom and in the number of surface atoms, and the work done may be used to define an effective surface stress tensor which is intermediate between the purely elastic and the completely liquid-like cases. Couchman *et al.* (1972) have discussed this complication in some detail. They find it possible to give a generalized discussion of the surface stress in terms of the linear components of the area per surface atom, and thus to avoid non-linear elastic terms which appear when the stress is defined in terms of the components of an area element of surface; this is analogous to the use of Eulerian rather than Lagrangian components in finite elasticity. However, this surface stress is not simply related to any readily defined surface strain, and the condition of zero surface strain can not be unambiguously defined.

Although complete equilibrium can seldom be attained, a polycrystalline mass at high temperatures will be able to reduce its grain boundary energy, and hence its total free energy, by atomic movements over relatively short distances. The grains will usually have originated from random centres, and there will thus be a statistical distribution of different polyhedral crystals. In order that these polyhedra shall completely fill space, certain purely mathematical conditions must be satisfied; when these are combined with the conditions for local surface

energy equilibrium, deductions may be made about the (metastable) equilibrium state, towards which the assembly will tend to move. This topological theory has been extensively investigated by C. S. Smith (1952).

It is instructive to consider first a two-dimensional analogy in which we have an assembly of polygons meeting along lines and at corners. We suppose that each separating edge possesses a certain free energy per unit length, and for simplicity we take this to be a constant, independent of the two polygons which it separates. The problem of minimizing the total free energy, subject to the restrictive condition that there are a given number of polygons present, reduces to that of finding the arrangement which gives minimum line (edge) length. Consider first corners which are the junctions of three edges. Then by considering small displacements of the edges, it is at once evident that the energy is minimized if the three edges make equal angles of  $120^\circ$  with each other. Similarly, any four-junction corner will be unstable, since a small re-adjustment will allow it to split into two three-junction corners of  $120^\circ$  as shown in Fig. 8.1. It follows that the arrangement of lowest energy is



FIG. 8.1. Unstable configuration (a) changes rapidly to (b) during two-dimensional growth.

one in which all corners are equi-angled three-edge junctions, and this is possible with straight edges if the whole two-dimensional structure consists of an array of regular hexagons. Such an array would be almost indefinitely metastable, since any displacement of any of the boundaries would result in an increase of energy.

We now suppose that our two-dimensional array has arisen by growth from random centres. This means that inevitably there will be some polygons with more than six edges and some with less. The approach to equilibrium will now take place in stages. Corners formed from four or more edges will still be very unstable and will dissociate into three-edge corners by small boundary displacements. Similarly, all three-edge corners will adjust their angles to the equilibrium  $120^\circ$  array. Since the polygons do not all have six edges, these adjustments mean that many of the edges must become curved. Now while a straight edge is indefinitely metastable in two dimensions, a curved edge is not. Provided material transfer across the boundary is possible, the edge will migrate slowly towards its centre of curvature, decreasing its area, and thus the total energy, as it does so. In a constrained two-dimensional assembly, where the edges meet at corners at  $120^\circ$ , not all of them will be able to migrate towards centres of curvature, but on the average this will be the dominating movement, and will result in the growth of some polygons at the expense of others. In this way, successive polygons may be eliminated entirely from the structure. The important conclusion is that this process will continue indefinitely until the whole assembly consists of polygons separated by straight boundaries passing from one edge of the assembly to the other, unless, by chance, the stable hexagonal array is achieved. Any actual random arrangement of polygons

thus tends to change, not into the metastable hexagonal array, but into a few very large polygons.

The two stages of the above process correspond to the attainment of local equilibrium over small regions of the boundaries, and of general equilibrium over the whole assembly. The whole process is a two-dimensional analogue of the grain growth of a metal, described in Part II, Chapter 19, and corresponds rather closely to actual observations of microstructural changes. However, the polygons in a microsection are sections of three-dimensional polyhedral grains, so we must further consider the equilibrium of a three-dimensional stack of such polyhedra.

We first consider the attainment of local surface energy equilibrium, which may be achieved at sufficiently high temperatures by the migration of atoms over short distances. Suppose the grain boundary surfaces of three grains of fixed orientations meet in a common line. Figure 8.2 shows a section through this line at a point  $O$ , where the direction of the line is

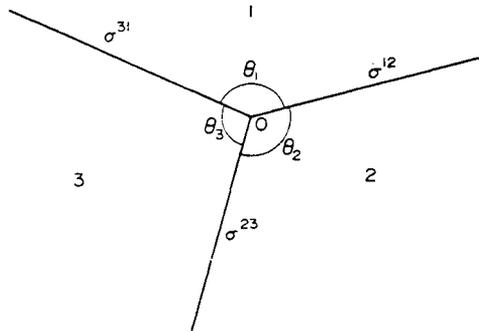


FIG. 8.2. Section through a three-grain junction.

normal to the plane of the paper. We assume that the grain boundary free energies,  $\sigma^{12}$ ,  $\sigma^{23}$ ,  $\sigma^{31}$  are constants determined by the orientations of the lattices concerned. The configuration attains local equilibrium when a virtual displacement of the boundary gives no first-order change in the free energy, and the condition for this is readily seen to be

$$\frac{\sigma^{23}}{\sin \theta_1} = \frac{\sigma^{31}}{\sin \theta_2} = \frac{\sigma^{12}}{\sin \theta_3}, \tag{35.1}$$

where  $\theta_i$  are the dihedral angles. In the special case where the surface stresses contain no shearing components, the surface free energies are numerically equal to tension forces in the boundaries, and the above equation gives the familiar triangle of forces condition. If the surface free energy is independent even of the lattice orientations, the free energies are all equal, and the dihedral angles are all  $120^\circ$ . As we shall see below, this is approximately true for large orientation differences.

The more general problem of the equilibrium of three such grains when the energies  $\sigma$  vary with the position of the boundary has been treated by Herring (1951). Figure 8.3 shows the same three grains as Fig. 8.2, and we consider a change resulting from the

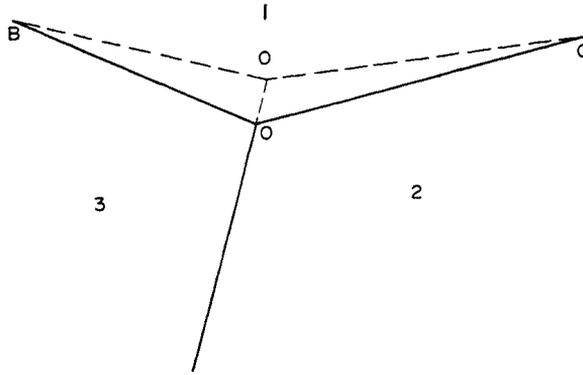


FIG. 8.3. To illustrate Herring's conditions for surface free energy equilibrium.

displacement of the line of intersection through  $O$  to a new position  $O'$  in the plane of the boundary between orientations 2 and 3. The other two boundaries are supposed to acquire slight kinks at  $B$  and  $C$ , where  $BO'$ ,  $CO'$  are much greater than  $OO'$ , but still infinitesimal. The first-order change of surface energy per unit length of line normal to the plane of the figure is then

$$\delta G = [\sigma^{23} - \sigma^{31} \cos(\pi - \theta_3) - \sigma^{12} \cos(\pi - \theta_2)] OO' - O'B(\partial\sigma^{31}/\partial\theta_3) \delta\theta_3 - O'C(\partial\sigma^{12}/\partial\theta_2) \delta\theta_2.$$

Since  $\delta\theta_3 = \widehat{BO'O} - \theta_3$ ,  $\delta\theta_2 = \widehat{CO'O} - \theta_2$ , we also have

$$O'B \delta\theta_3 = O'O \sin \theta_3, \quad O'C \delta\theta_2 = O'O \sin \theta_2$$

and equating  $\delta G = 0$ ,

$$\sigma^{23} + \sigma^{31} \cos \theta_3 + \sigma^{12} \cos \theta_2 - (\sin \theta_3) (\partial\sigma^{31}/\partial\theta_3) - (\sin \theta_2) (\partial\sigma^{12}/\partial\theta_2) = 0. \quad (35.2)$$

In this equation, the derivatives  $\partial\sigma^{31}/\partial\theta_3$ ,  $\partial\sigma^{12}/\partial\theta_2$  are measured in opposite directions of rotation. If derivatives with respect to  $\theta$  are evaluated in the same sense as the labelling of the phases (in this case in the clockwise direction of rotation), we may write alternatively

$$\sigma^{23} + \sigma^{31} \cos \theta_3 + \sigma^{12} \cos \theta_2 - (\sin \theta_3) (\partial\sigma^{31}/\partial\theta) + (\sin \theta_2) (\partial\sigma^{12}/\partial\theta) = 0. \quad (35.3)$$

Equations (35.2) or (35.3), together with two similar relations which may be written down by inspection, are Herring's conditions for equilibrium when three grains meet along a line. The set of equations, like the simpler eqns. (35.1), contains only two independent relations. If the boundary free energies are constants, there are only three unknowns, and experimental determinations of the dihedral angles enable the ratios of the three energies to be determined. In the general case, there are five unknown ratios of the energies and their derivatives, so that dihedral angle measurements are unable to establish the grain boundary parameters.

The arguments used for the two-dimensional array may now be extended. Again, a random array will never achieve a metastable equilibrium, since the requirements of local equilibrium will inevitably result in grains with curved faces. In principle, such grains will migrate under

surface energy forces until only a few large grains are left, although the variation of boundary energy with orientation may introduce some modifications in the simple picture. Thus a few boundaries (coherent twin boundaries) may have energies which vary so rapidly with orientation of the boundary surface that virtually no movement of the surface is possible, and these will be frozen into the structure indefinitely. Most boundaries, however, may be supposed to become mobile at sufficiently high temperatures, so that grain growth should always occur.

In two dimensions, there is a theoretical possibility of achieving a stable array by regular packing. This is no longer true in three dimensions. In the simplest example (constant grain boundary energy), the requirements of local surface energy equilibrium are that all surfaces should meet each other in groups of three at angles of  $120^\circ$  along lines which themselves meet in groups of four at angles of  $\cos^{-1}(-\frac{1}{3}) = 109^\circ 28'$ . As in two dimensions, junctions at which four or more grains meet along a line will rapidly separate into two junctions of three grains. However, there is no regular polyhedron with plane faces having edges meeting at the correct angle, and although a pentagonal dodecahedron is a close approximation, this figure lacks the symmetry required to form the basis of a repeating pattern filling all space. The problem was considered long ago by Lord Kelvin, who showed that if a b.c.c. stack of regular tetrakaidecahedra (truncated octahedra), shown in Fig. 8.4, is modified

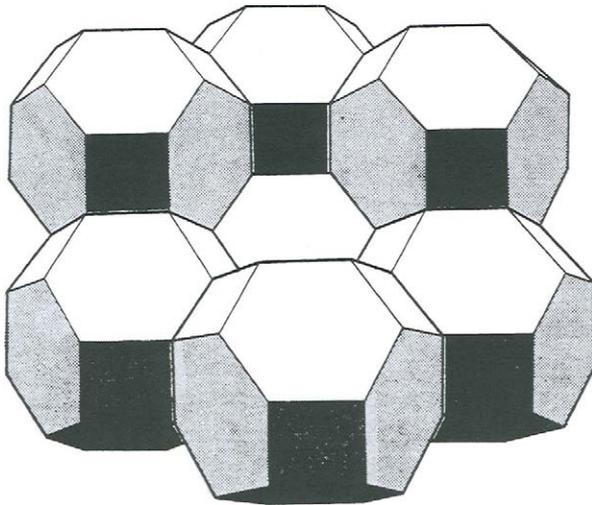


FIG. 8.4. A stack of regular tetrakaidecahedra (after Smith, 1952).

by introducing double curvatures into the hexagonal faces, the local surface tension requirements can all be satisfied. The curvatures, as we have pointed out, prevent the array from achieving a stable configuration.

Attempts to find an equilibrium stack require a combination of surface energy conditions with the purely mathematical conditions which ensure that the stack shall fill all space. In two dimensions, the maximum number of corners per polygon cannot exceed two. An array of uniform hexagons, which possesses this maximum sharing of corners, also gives the

shortest edge length for a given number of cells. Attempts were made by Smith (1952) to invoke similar topological principles for three-dimensional arrays. He stated that the maximum number of corners per polyhedron in a space filling array is six, and that minimum interfacial area occurs when there is a maximum sharing of faces and corners. These statements were retracted in a later paper (1953), where processes which lead to an apparently limitless increase in the number of corners, edges and faces per polyhedron were described. This later paper also withdraws a former "proof" that for minimum total interfacial area, the average number of corners on each face is  $5\frac{1}{2}$ , although reasons for believing that this statement is correct are given. In a later note, Meijering (1953) showed that the number of corners per polyhedron can exceed six, even when all faces are plane. Purely topological arguments thus do not seem to lead to the same sort of useful result in three dimensions as they do in two dimensions, although many interesting relations are pointed out in Smith's papers.

In discussing problems which depend upon grain size, it is often useful to have expressions for the mean boundary area, mean edge length, and mean number of corners per unit volume in terms of the mean grain diameter. Probably the most realistic simple assumption is to consider all grains to be equal tetrakaidecahedra, as above. If the separation of square faces is  $L^B$ , the edge length is  $L^B/2\sqrt{2}$  and the number of grains per unit volume is  $2/(L^B)^3$ . The grain boundary area,  $vO^B$ , grain edge length,  $vL^E$ , and number of grain corners,  $vN^C$ , all per unit volume, are then given by

$$vO^B = 3.35/L^B, \quad vL^E = 8.5/(L^B)^2, \quad vN^C = 12/(L^B)^3. \quad (35.4)$$

In a real assembly, these equations will be approximately valid for some mean grain diameter  $L^B$ . Although the numerical constants will change, the functional dependence will be correct for grains of any shape.

### 36. DISLOCATION MODELS OF GRAIN BOUNDARIES

Consider the following process, leading to a general grain boundary. Two pieces of crystal are each cut along a plane, so that the two cut planes correspond to the orientations of the required boundary relative to the crystals it separates. This uses up four degrees of freedom. The boundary is now made by cementing the crystals together along the cut planes, the remaining degree of freedom being used to specify the azimuthal orientation about the normal to these planes. If the atoms are held rigidly in position, it is obvious that the atomic bonding across the boundary will be quite different from that in the interior of the crystals. The changes in configuration and in interatomic distances will result in the boundary energy per atom being much larger than that in the interior of the crystals, and the whole of this extra misfit energy may be considered to be localized in the immediate vicinity of the interface. In many cases, however, the misfit may be reduced considerably by readjustment in the positions of the atoms near the boundary. These readjustments, in turn, disturb the ideal arrangement in regions a little further from the boundary, and thus cause further small atomic displacements. The displacements evidently fall off rapidly with distance from the boundary, and at sufficient distances are equivalent to elastic strains of

the two lattices. The boundary energy now contains two terms, a misfit energy localized at the boundary, and an elastic energy which in principle extends throughout the crystal but which is largely concentrated in a small volume near the boundary. This illustrates the importance of the general definition of boundary energy given in the last section.

The actual atomic configuration and the relative importance of the two terms will be such as to minimize the total boundary free energy. In general, the greater the amount of misfit when rigid structures are joined, the less effective will be readjustments of atomic position in lowering the energy. Thus grain boundaries separating lattices with large relative orientation have high energies localized at the interface, whereas boundaries separating lattices of nearly the same orientation have lower energies which are mainly spread through a larger volume. We have the apparently anomalous result that the smaller the misorientation, the greater is the distance over which atomic adjustments are made, and hence the "wider" is the boundary. For a small misorientation, the boundary region does not correspond to uniform disregistry of the atoms on the two sides, but is rather to be pictured as a surface over which most atoms are in almost perfect registry, and a few atoms are not in registry. This is a special example of a type of boundary in the solid state which we shall describe as semi-coherent. The centres of strain are dislocations, and a surface array of dislocation lines forms a suitable model for such a boundary.

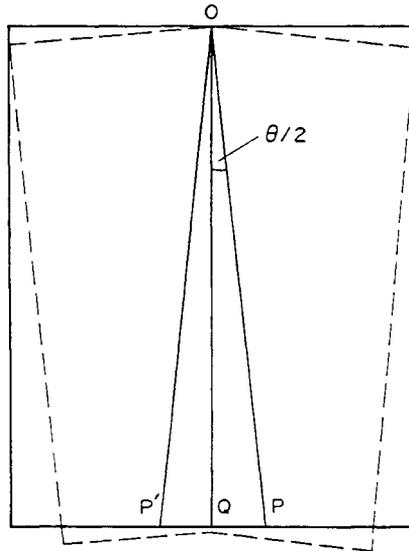


FIG. 8.5. The formation of a tilt boundary.

In the simplest type of interface, the two lattices are connected by a rotation about an axis lying in the boundary. This may be constructed macroscopically by removing a wedge of material  $POP'$  (Fig. 8.5) from a perfect crystal and rejoining along  $OP$ ,  $OP'$ . When  $\widehat{OPP'} = \widehat{OP'P}$ , the joining surface is equally inclined to the two lattices, and we have a symmet-

rical tilt boundary. A dislocation model for this configuration was first suggested by Burgers (1939) and Bragg (1940). On an atomic scale, the surfaces  $OP, OP'$  are not plane but are stepped, as shown in Fig. 8.6(a). Each step corresponds to the ending of a plane of atoms, or more generally to a lattice plane, now running towards the boundary from left or right, but originally parallel to the  $x_2$  axis. By elastic distortion of the crystals, the two surfaces may be fitted together as shown in Fig. 8.6(b). Most atoms are then in registry except near the steps. As the steps are the edges of incomplete atomic planes, they may be considered

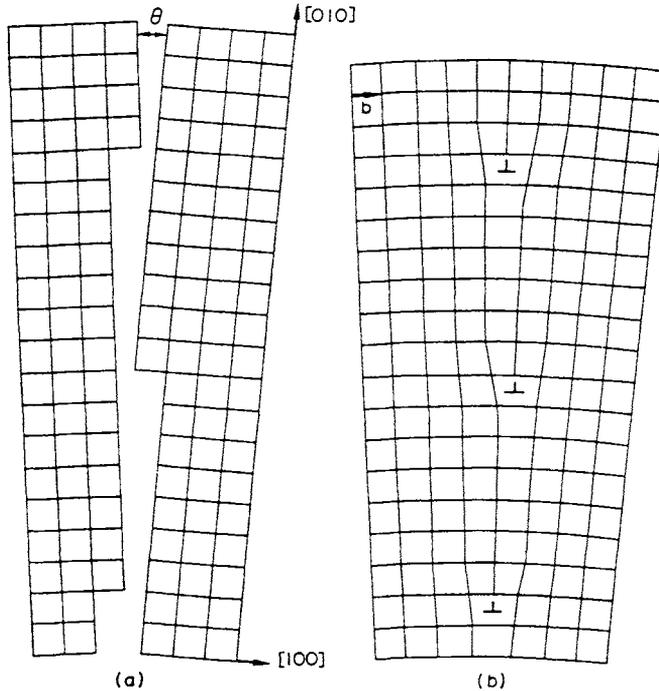


FIG. 8.6. Dislocation structure of a symmetrical low angle tilt boundary in a simple cubic structure (after Read and Shockley, 1950).

to be edge dislocations in the boundary. The example shown may be visualized as a simple cubic lattice, with the  $x_3$  (or  $\langle 001 \rangle$ ) axis as the axis of rotation, the plane  $x_1 = 0$  as the boundary, and the Burgers vector of the dislocations along the  $\langle 100 \rangle$  direction of one of the lattices. Equally, the lines may be interpreted as giving the strain pattern of a symmetrical tilt boundary in any structure. From the figure, we see that if the spacing of the dislocation lines (measured along the boundary) is  $Y$ , and the Burgers vector of each is  $b$ , the total rotation between the lattices is

$$\theta = 2 \sin^{-1}(b/2Y) \approx b/Y, \quad (36.1)$$

where the approximation is valid only for small tilts.

We notice that although the steps were assumed to belong to one or other half-crystal, the edge dislocations are present in the boundary itself rather than in one or other lattice. The quantity  $b$  in eqn. (36.1) is the magnitude of the Burgers vector defined in a reference

crystal in the usual manner, and the equation is independent of the orientation of this reference crystal. Dislocations in boundaries clearly differ in some important respects from dislocations in the interior of crystals, one obvious difference being that the Burgers vector of a boundary dislocation is not fixed in space. In discussing the definition of the Burgers vector, we have always tacitly assumed that the reference crystal has the same mean orientation as the real crystal. For a dislocation in a boundary, we may choose the reference crystal to correspond in orientation to either of the real crystals, or to some intermediate situation. In principle, the boundary dislocation may move into either lattice, the Burgers vector appropriate to this movement being defined by using a reference crystal in the orientation of the lattice into which it moves. The local Burgers vector suffers a sharp discontinuity as we pass through the boundary. These geometrical relations are obvious, but are emphasized here because care is sometimes needed before dislocations in boundaries can be assigned the properties normally associated with lattice dislocations.

Equation (36.1) may also be derived from Fig. 8.5 since  $PP'$  is the total Burgers vector of the dislocations in a length  $OP$  of boundary. The equation is sometimes presented in the alternative form

$$\theta = 2 \tan^{-1}(b/2Y'), \quad (36.1a)$$

where  $Y'$  is the dislocation spacing measured normal to the slip planes in one of the crystal lattices, i.e. parallel to  $OQ$  in Fig. 8.5 rather than to the interface  $OP$ . Although the dislocations are equally spaced in Fig. 8.6, this cannot be arranged for arbitrary  $\theta$  since the actual separation of dislocations parallel to  $OQ$  must be an integral multiple of the interplanar spacing  $d$ . There is thus generally some variability in the individual separations, and eqns. (36.1) and (36.1a) refer to mean values  $Y, Y'$  of these separations. Uniform spacing is possible only at discrete values of  $\theta$ , where  $(2 \tan \frac{1}{2}\theta)^{-1}$  is an integral multiple of  $d/b$ , or for small misorientations when  $\theta = b/nd$ , where  $n$  is an integer. According to the strict definition of a twin, all uniform spacings correspond to coherent twin boundaries inasmuch as the boundary plane must be rational and the lattices on each side are mirror images (Brooks, 1952). However, the low angle boundaries of this type have very high indices, and would not be classed as twins in normal circumstances. The more important twin boundaries, with low indices, correspond to small integral values of  $Y'/d$ ; they are often represented as coincidence site boundaries (see below).

A tilt boundary may be formed in an originally single but imperfect crystal if a number of edge dislocations of the same sign move by glide and/or climb processes until they are aligned parallel to each other and normal to the slip planes. As each dislocation moves, the lattice and shape deformations (p. 322) change, and when the dislocations are aligned, the lattice dislocations have added together so as to produce a small rotation. We recall that an isolated edge dislocation in a finite crystal produces a rotation of the lattice planes (p. 274), this being the tilt boundary of minimum angle. The shape deformation produced by the lining up of the dislocations is discussed below; the array of edge dislocations provides one of the simplest examples of the compatibility conditions discussed on p. 319.

Once formed, the tilt boundary is stable against dissociation by loss of individual dislocations into the crystals on each side. This is because the force between parallel dislocations

on parallel, non-identical, glide planes pulls them into alignment (eqn. 30.26). At distances from the boundary greater than the mean dislocation separation, the effects of the dislocations will be nearly identical with those produced by removing a uniform wedge. The stress field of the dislocation array thus extends only for an effective distance of the order of  $Y$  into the crystals, since the displacements cancel at larger distances. The expression for the energy of each dislocation in a small angle boundary will thus contain a term in  $\ln Y$  to replace the  $\ln r_c$  of eqn. (30.15). Note that the smaller the misorientation the greater the extent of the elastic field, as we remarked above. However, when the dislocation spacing is non-uniform, there are additional, long-range terms in the expression for the elastic energy.

Equation (36.1) in its exact form is valid in principle even for large misorientations. As  $\theta$  increases, the dislocations become closer together, until eventually  $Y$  is of the order of the core radius,  $r_c$ . It is clear that the boundary must be considered as a whole for separations much smaller than this, and its analysis in terms of individual dislocations has only formal significance. The dislocation interpretation is nevertheless still extremely useful in considering compatibility relations, since continuous distributions may be assumed; the physical structure of such boundaries is considered later in this section. Lomer and Nye (1952) used the bubble model to study the change in boundary structure as  $\theta$  increases; this is of value since the model gives a reasonable approximation to the interatomic forces in a metal like copper. They found that when the spacing of the dislocations becomes of the order of their original widths (see p. 274), this width begins to decrease. Eventually, the dislocations coalesce, so that the defects resemble closely spaced vacancies, and the whole boundary is incoherent.

The symmetrical tilt boundary has only one degree of freedom. An extra degree of freedom is obtained if we allow the boundary to rotate about the direction common to the two lattices, i.e. if in Fig. 8.5,  $\widehat{OPP'} \neq \widehat{OP'P}$ . The crystal spacings along the  $x_2$  direction are no longer equal, and some of the  $\{100\}$  planes must also end in the boundary. The asymmetrical tilt boundary thus contains dislocations of two types, as shown in Fig. 8.7. The angles which the boundary plane makes with the  $[010]$  directions are now  $\frac{1}{2}\theta + \varphi$  and  $\frac{1}{2}\theta - \varphi$ , where  $\varphi = 0$  for the symmetrical boundary. By considering the number of  $\{010\}$  planes which end on unit length of the boundary from the two sides, it is readily seen that for small misorientations the spacing of  $[100]$  dislocations is given by

$$Y_{100} = b/\theta \cos \varphi. \quad (36.2)$$

Similarly, the spacing of the  $[010]$  dislocations is obtained by finding the number of  $\{100\}$  planes which approach the boundary from the two sides, and is

$$Y_{010} = b/\theta \sin \varphi. \quad (36.3)$$

When  $\varphi = 0$ , only the  $[100]$  set are present; when  $\varphi = 90^\circ$ , only the  $[010]$  set. The latter case represents a symmetrical tilt boundary with axis of rotation again  $[001]$ , but the boundary surface perpendicular to  $[010]$ .

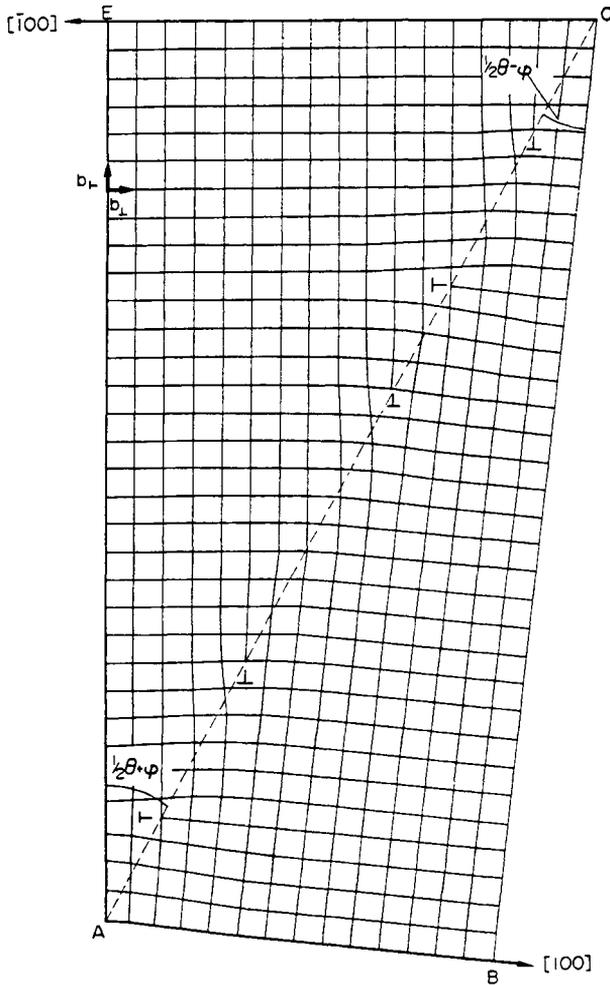
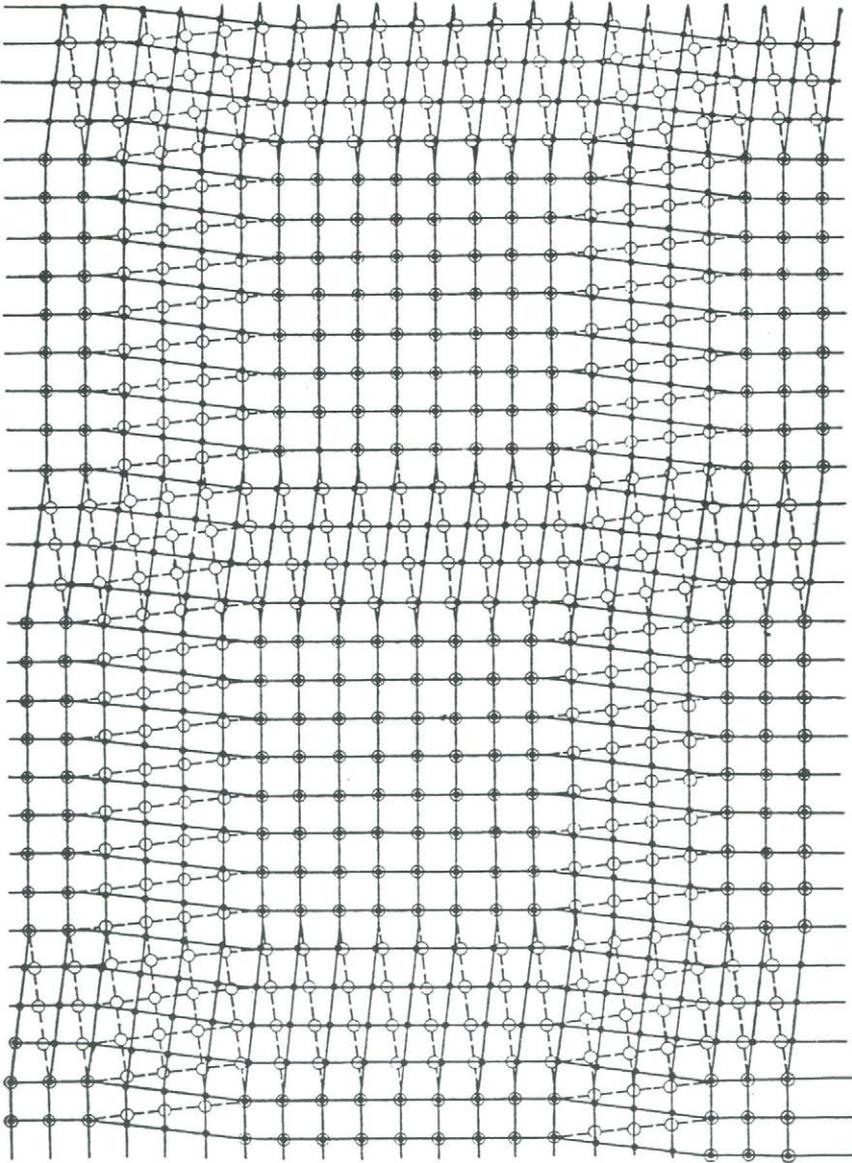


FIG. 8.7. Structure of an asymmetrical tilt boundary (after Read, 1953).

In the examples we have been considering, the tilt boundaries have been constructed from edge dislocations alone. It will only be possible to do this with allowable lattice dislocations, however, when the axis of rotation is perpendicular to a possible Burgers vector, or to two such vectors for an asymmetric boundary. In other cases, the boundary could be constructed of straight dislocation lines of mixed edge and screw type parallel to the axis of rotation. The Burgers vectors of these dislocations would have to be such that the edge components are equal and give the required tilt, while the screw components alternate in sign, and do not contribute to the misorientation. The existence of the screw component does, however, increase the boundary free energy.

We next consider how further degrees of freedom in the relative orientation of the two lattices may be obtained. In Fig. 8.6 we may give the two crystals an independent rotation about the  $x_2$  axis by inserting a second set of dislocations, perpendicular to the first, but



**FIG. 8.8.** Structure of a twist boundary (after Read, 1953). Open and filled circles represent atoms on each side of the boundary which is parallel to the plane of the figure.

having the same Burgers vector. For the most general relation, we also require a rotation about  $x_1$ , i.e. about an axis perpendicular to the boundary, and we shall now show that this may be achieved by arrays of screw dislocations. When the only misorientation is a relative rotation about this normal, we have a simple twist boundary.

In the simple tilt boundary, atomic planes from the two crystals meet the boundary along parallel lines but with different densities. In a twist boundary, the densities of atomic planes are equal on both sides, but corresponding planes do not meet in parallel directions. The function of the screw dislocations is to provide this change of direction whilst allowing most atoms at the interface to remain in registry. In Fig. 8.8 we show a  $\{100\}$  twist boundary in a simple cubic structure. The boundary is in the plane of the figure, open circles representing the atoms on one side of the boundary, and filled circles the atoms on the other side. The  $[010]$  and  $[001]$  rows of atoms have different mean directions on the two sides of the boundary, but the introduction of the screw dislocations enables these rows at the interface to remain parallel over most of their lengths. The relative rotation of the  $[010]$  rows, for example, is corrected by a set of parallel screw dislocations of spacing  $Y = \frac{1}{2}b \operatorname{cosec}(\theta/2)$ , where  $\theta$  is the angle of twist between the two lattices. This set of screw dislocations alone would produce a long-range shear in the lattices, and the  $[001]$  rows on the two sides of the boundary would still not be parallel. The insertion of a second set of screw dislocations, perpendicular to the first and having the same spacing, brings the  $[001]$  rows into coincidence over most of the boundary, and removes the long-range shear. The stress field now extends only for a distance of order  $Y$  from the boundary, and the  $[100]$  rows of atoms are continuous across the interface (except near the dislocations). A crossed grid of screw dislocations thus gives a stable twist boundary, the misorientation again being given by eqn. (36.1).

Twist boundaries can only be made from pure screw dislocations if the boundary surface contains two perpendicular directions corresponding to allowable Burgers vectors, as does, for example, a f.c.c.  $\{100\}$  boundary. In general, dislocations of mixed type must be used, the edge components being arranged to cancel one another. In the f.c.c. structure, a simple hexagonal net of slip dislocations will give a twist boundary on a  $\{111\}$  plane (Frank, 1955). There are thus many possibilities for low angle dislocation boundaries in real structures, but most of them are based on simple rectangular or hexagonal nets. Slight deviations in the orientation of the boundary plane, or of the orientation relations from the simplest types, are then achieved by having local "faults" in the mesh.

The simple boundaries are specified by only one or two parameters, and it is usually easy to find the correct dislocation description. Models of more general boundaries will clearly involve rather complex arrays of dislocations of several Burgers vectors. Even for small misorientations, the representation of such a boundary by a particular array is not unique, and it is not immediately obvious which of the possible descriptions best corresponds to the actual configuration. Experimental studies of the dislocation structure of many low angle boundaries have been made, using precipitation techniques in transparent crystals, or thin film electron microscopy in metals. The boundaries usually approximate to the simple symmetrical types (for details see, e.g., Frank, 1955; Amelinckx, 1957, 1958; Carrington *et al.* 1960), but we shall not discuss them here.

For a general boundary we may specify a resultant or net dislocation content, as first shown by Frank (1950). The following method for finding this net content is equivalent to that originally given by Frank, except that we use rotation matrices instead of vector notation. As we shall see in the next section, Frank's formula is a particular case of a more general expression for the dislocation content of a boundary between two crystals of different structure and orientation.

Consider a reference lattice with origin  $O$ . Two crystals of different orientations may be constructed from this reference lattice by giving it arbitrary rotations represented by  $\mathbf{R}_+$  and  $\mathbf{R}_-$  respectively. We consider these two crystals to meet along a boundary surface with normal  $\mathbf{v}$ , and we now show how this configuration can be achieved by introducing dislocations into the surface  $\mathbf{v}$ .

Let  $OP = \mathbf{p}$  be a large vector in the boundary, and consider a right-handed Burgers circuit  $PA_+OA_-P$  (Fig. 8.9(a)) where  $A_+$  is any point of the  $+$  lattice and  $A_-$  is any point

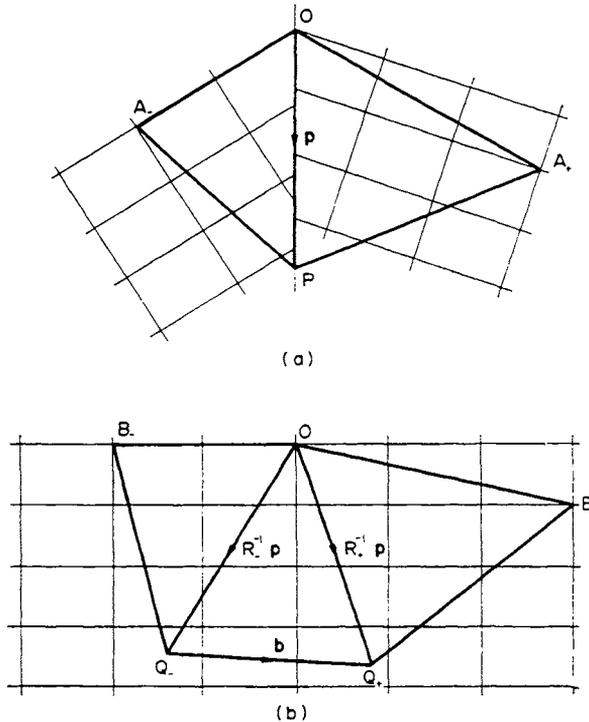


FIG. 8.9. To illustrate Frank's procedure for defining the net Burgers vector of the boundary dislocations intersecting a vector  $\mathbf{p} = OP$  in a general grain boundary.

of the  $-$  lattice. The corresponding path  $Q_+B_+OB_-Q_-$  in the reference lattice is obtained by applying the inverse rotations  $\mathbf{R}_+^{-1}$  and  $\mathbf{R}_-^{-1}$  to the parts  $PA_+O$  and  $OA_-P$  of the circuit respectively. This path is shown in Fig. 8.9(b) and the closure failure  $Q_-Q_+ = OQ_+ - OQ_- = (\mathbf{R}_+^{-1} - \mathbf{R}_-^{-1})\mathbf{p}$ . Thus with the sign convention of p. 245, the net Burgers vector is given by

$$\mathbf{b} = (\mathbf{R}_+^{-1} - \mathbf{R}_-^{-1})\mathbf{p}. \quad (36.4)$$

Equation (36.4) gives the total Burgers vector of the dislocations in the surface crossing any vector  $\mathbf{p}$  in the surface, referred to the reference lattice specified above. Frank gave the equivalent vector expression, taking  $\mathbf{R}_+$  to represent a rotation of  $\theta_+$  about a unit vector  $\mathbf{l}_+$ , and  $\mathbf{R}_-$  to represent a rotation of  $\theta_-$  about a unit vector  $\mathbf{l}_-$ . This is

$$\mathbf{b} = (\cos \theta_- - \cos \theta_+) \mathbf{p} + \mathbf{p} \wedge (\mathbf{l}_+ \sin \theta_+ - \mathbf{l}_- \sin \theta_-) + (\mathbf{l}_+ \cdot \mathbf{p}) (1 - \cos \theta_+) \mathbf{l}_+ - (\mathbf{l}_- \cdot \mathbf{p}) (1 - \cos \theta_-) \mathbf{l}_-. \quad (36.5)$$

It will be noted that the procedure we have used to deduce eqn. (36.4) is essentially identical with that used on p. 319 for finding the dislocation tensor associated with a continuous distribution of dislocations. We may obtain eqn. (36.4) in fact by assuming that we have a continuous distribution of dislocations confined to a thin sheet (Bilby, 1955), as discussed in Section 38.

The net dislocation content is a geometrical property of the boundary considered, but it is not a unique description since the relations between the two lattices can be expressed in an infinity of different ways, each of which leads to a different  $\mathbf{b}$ . This problem is considered below; for the present we note that for low angle boundaries a particular description has physical significance and  $\mathbf{b}$  in eqn. (36.4) may then be considered to be the sum of the contributions from the individual dislocations forming the boundary. We have seen that in three dimensions the appropriate dislocation density is given by the number of dislocations of any type crossing a unit area of any orientation. Similarly, in a two-dimensional distribution an appropriate density is the number of dislocations crossing unit length of line of any orientation in the boundary. Suppose now that  $\mathbf{p}$  is a unit vector. Then if  $n_i$  dislocations, each having Burgers vector  $\mathbf{b}_i$ , cross  $\mathbf{p}$ , the net Burgers vector  $\mathbf{b}$  is given by the sum

$$\mathbf{b} = n_i \mathbf{b}_i. \quad (36.6)$$

The number of possible terms on the right-hand side of eqn. (36.6) is restricted to the number of stable dislocations of the lattice, at least in cases where a sub-boundary is formed physically by amalgamation of lattice dislocations. If there are only three non-coplanar Burgers vectors to be considered, the resolution of  $\mathbf{b}$  into its components is unique. When, as is more usual, there are more than three Burgers vectors of lowest energy,  $\mathbf{b}$  may be compounded into component dislocations in several ways. Since  $\mathbf{b}$  is directly proportional to  $\mathbf{p}$ , and  $\mathbf{b}_i$  is fixed,  $n_i$  must be proportional to  $|\mathbf{p}|$ . The dislocations of type  $i$  are thus (statistically) uniformly spaced along any direction  $\mathbf{p}$  in the boundary, and must all be straight lines. Determination of  $n_i$  for two different boundary directions uniquely specifies the direction and spacing of the dislocation lines with Burgers vector  $\mathbf{b}_i$ . Thus any grain boundary may be defined in a reference lattice by arrays of dislocation lines, having as a minimum three different sets of non-coplanar vectors. When several different models are possible, they represent different ways of making the atomic transition from one grain to the other compatible, and they are no longer distinguishable when the dislocation distribution is regarded as continuous. Nevertheless, the lowest energy will generally correspond to the model in which the (discrete) dislocation density is lowest. The above remarks apply in

principle to all boundaries, but as already remarked, the dislocation model begins to lose its physical significance when the relative orientations of the grains becomes large.

We have not said anything about the choice of reference lattice orientation in which the Burgers vectors are defined. The most convenient reference lattice is usually in an orientation equivalent to that of one of the crystals, or else is the median lattice from which the two real lattices are obtained by equal and opposite rotations. The first choice is particularly useful when we consider the motion of an interface by glide motions of the dislocations, as in the theory of martensite or the stress induced migration of a symmetrical low angle tilt boundary. For example, if we use the  $(-)$  lattice as the reference crystal, we have, from (36.4),

$$\mathbf{b} = (\mathbf{R}^{-1} - \mathbf{I})\mathbf{p}, \quad (36.7)$$

where  $\mathbf{R}$  is the rotation giving the relative orientations of the two crystals.

When the median lattice is used as the reference lattice, the dislocation density takes a particularly simple form. The two real lattices are now generated by rotations  $\theta_+ = -\theta_- = \frac{1}{2}\theta$  about a common direction  $\mathbf{l}$ , and eqn. (36.5) becomes

$$\mathbf{b} = 2 \sin \frac{1}{2}\theta \mathbf{p} \wedge \mathbf{l}. \quad (36.8)$$

The more general expression (36.4) or (36.5) need be retained only when it is desired to have a common reference lattice for a number of different boundaries. This occasionally has some advantages; for example, the total Burgers vectors are conserved at a "surface node" where three or more boundaries meet.

We now consider the atomic structures of high-angle grain boundaries where the above dislocation analysis has only formal significance and the misfit is concentrated mainly in the immediate vicinity of the boundary, which has a width of only a few atoms. One possible assumption is that there is practically no correlation of atomic positions across the boundary, and that the positions of the atoms in the boundary region may only be described statistically as in a two-dimensional disordered or liquid-like array. However, there is now strong evidence, both from structural studies and from measurements of physical properties such as grain boundary diffusion, that high-angle boundaries have some form of periodic structure. An early suggestion of this type (Mott, 1948) was that the boundary structure consists of "islands" of good atomic matching, separated by regions where the matching is poor. The islands are small groups of atoms, and the disordered regions have a structure rather like that of a liquid. This is similar to a dislocation model except that the bad matching regions cannot be pictured as isolated dislocations. As the misorientation of the grains for a boundary of given type is increased, the structure of the boundary might be expected to change from the low-angle dislocation model to Mott's model, and eventually to a completely disordered boundary. Such a proposal was actually made by Smoluchowski (1952) for tilt boundaries in connection with some early results on grain boundary diffusion. The misfit regions were assumed to be formed originally by the coalescence of groups of individual regions when the tilt angle exceeds  $\sim 15^\circ$ .

Smoluchowski's proposal is similar to some recent descriptions of grain boundaries in terms of disclinations rather than dislocations (e.g. Li, 1972). As pointed out in Section 30,

a disclination of arbitrary strength is topologically equivalent to the internal termination of a grain boundary, and in particular a wedge disclination represents the termination of a tilt boundary. It follows that a wedge disclination dipole is formally equivalent to an edge dislocation when the separation of the opposite disclinations is of the order of the atomic distance, and to a finite wall of edge dislocations when the separation is larger. Now consider a symmetrical tilt boundary formed by edge dislocations and suppose that as the tilt angle and hence the density of dislocations increase the dislocations begin to cluster into closely spaced walls separated by larger distances. When the walls become very dense, the individual dislocations lose their identities and the walls may be considered to have condensed into disclination dipoles (Mott's bad regions) separated by regions of good fit. The relation between the usual dislocation model and this disclination model is shown in Fig. 8.10. In the figure, each disclination dipole replaces a single dislocation, so that the

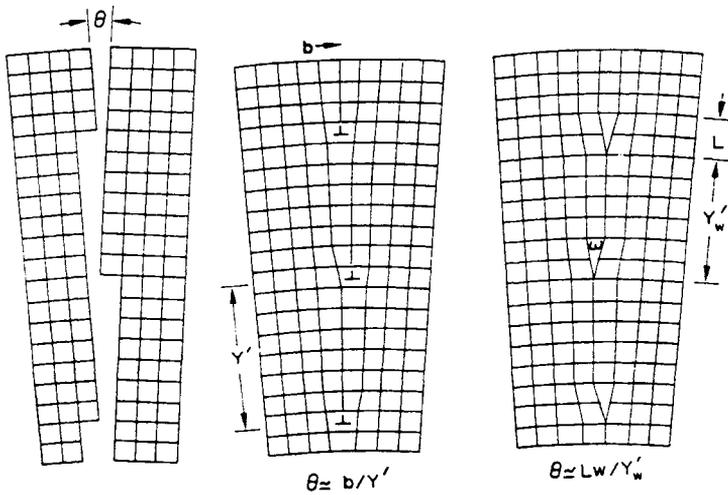


FIG. 8.10. Alternative models of a symmetrical low angle tilt boundary as a wall of edge dislocations of Burgers vector  $b$  and spacing  $Y'$  or as a wall of wedge disclination dipoles of strength  $\omega$ , element separation  $L$ , and spacing  $Y'_w$  (after Li, 1972).

spacings  $Y'$  and  $Y'_w$  are equal, but the disclination model is considered to be more appropriate to the high-angle situation where  $Y'$  becomes small but  $Y'_w$  remains appreciably larger than  $b$ .

Other recent models of grain boundary structure also embody Mott's concept, but are differently formulated. These theories arose from the discovery that certain high-angle boundaries have special properties which suggest that their structures are simpler and their energies lower than random boundaries. The most obvious example of a special boundary is a coherent twin interface, but there are many other boundaries of unusually good atomic fit. These boundaries appear to have higher mobilities than random boundaries, at least when impurities are present in solution, as will be fully discussed in Part II, Chapter 19. In discussing results on secondary recrystallization, Kronberg and Wilson (1949) plotted the positions of the atoms in the two grains that meet at a boundary of high mobility. They

found that if the lattices are assumed to interpenetrate, an appreciable fraction of the lattice sites are common to both structures, and thus form a "coincidence site lattice". The concept of the coincidence lattice has had an important influence in the development of theories of high-angle boundaries.

Let two lattices interpenetrate to fill all space, and assume that there is a common lattice point. In general, there may then be other lattice points which coincide, and the set of such points forms the coincidence site lattice. In some special situations, the reciprocal density  $\Sigma$  of coincidence sites relative to ordinary lattice sites may be small, but for an arbitrary orientation  $\Sigma$  may tend to infinity and there is then effectively no coincidence site lattice. A model of a grain boundary is introduced by fixing the interface somewhere in the space and placing atoms on the appropriate sites of one lattice on one side of the interface and of the other lattice on the opposite side of the interface. If this boundary contains a high density of sites of the coincidence lattice, representing regions of good fit, the structure of the boundary should have the periodicity of a planar net of the coincidence site lattice. This does not mean that the model gives the actual atomic positions; if these are now allowed to relax to a minimum energy configuration, the further displacements may include a translation by some fraction of a lattice vector of the atoms at infinity on one side of the boundary with respect to those at infinity on the other, so that there are no real coincidences. Nevertheless, boundaries which contain high densities of coincidence sites before relaxation may be expected to represent better fit between the lattices, and thus to have low energies relative to other boundaries.

The coincidence site lattice depends only on the relative orientation of the two grains and is independent of the orientation of the interface, but the periodicity of the boundary structure will, of course, vary with the interface orientation. Low values of  $\Sigma$  result from grains rotated with respect to each other about a common axis which is normal to a close-packed plane, and the actual conditions are rather restrictive. If the two lattices are related by a transformation  $\mathbf{S}$ , the lattice points  $\mathbf{w}$  of the first lattice which are also points of the coincidence lattice must satisfy

$$w_i = S_{ij}u_j, \quad (36.9)$$

where  $\mathbf{u}$  is also a lattice point so that  $w_i$  and  $u_i$  are all integers. In general solutions do not exist unless  $\mathbf{S}$  can be expressed as a pure rotation  $\mathbf{R}$ . The simpler coincidence site lattices formed from two cubic lattices differing by a rotation have been listed by several authors, and Ranganathan (1966) has developed a systematic procedure for deriving these lattices by means of a generating function.

Consider a planar rectangular net of axial ratio  $\gamma$  which is rotated about the normal to the plane which passes through a lattice point of the net. If the net is used to define an orthogonal coordinate system, each rotation of a point with coordinates  $x, -y$  into a new position  $x, y$  will give rise to a coincidence lattice formed from the original and rotated nets. The angle of this rotation is

$$\theta = 2 \tan^{-1}(\gamma y/x) \quad (36.10)$$

and the area ratio of the two-dimensional cell of the coincidence and original lattices is

$$\Sigma' = x^2 + \gamma^2 y^2. \quad (36.11)$$

The lattice sites in any plane of a cubic crystal may be represented as a rectangular net of axial ratio  $\gamma = (h_i h_i)^{1/2}$ , or as a series of interpenetrating nets of this type, and it follows that the above function (36.11) can be used to generate the coincidence site lattices for rotations about the normal to the plane  $\mathbf{h}$ . However, it may be necessary to verify by inspection that all lattice sites are brought into coincidence by this operation, since otherwise eqn. (36.11) gives a sub-multiple of the correct reciprocal density of coincidence sites, and also that there are no additional coincidences, in which case it gives a multiple. Additional coincidences always occur if  $\Sigma'$  is even, and the true density of coincidence sites is then twice, or more generally  $2^n$  times, that given by eqn. (36.11). Otherwise, Ranganathan states that the factors leading to the above possibilities tend to cancel, so that the equation usually gives the true multiplicity. Thus all that is necessary in most cases is to assign relatively prime, integral values of  $x$  and  $y$  in the generating function  $x^2 + h_i h_i y^2$  and to divide even values of  $\Sigma'$  repeatedly by two until an odd value corresponding to  $\Sigma$  is obtained.

Obviously when some coincidence lattice relations are known, others may be generated by combining these together. Thus with two rotations  $\theta_1$  and  $\theta_2$  about the same axis corresponding to relatively prime reciprocal densities  $\Sigma_1$  and  $\Sigma_2$ , a new lattice of reciprocal density  $\Sigma_1 \Sigma_2$  will represent rotations of  $\theta_1 \pm \theta_2$  about this axis. It should also be noted that any orientation represented by an axis-angle pair  $\mathbf{l}, \theta$  can be represented in a number of different ways by making use of the symmetry properties of the lattice. Thus for any rotation matrix  $\mathbf{R}$  there are in the cubic system up to 23 non-equivalent alternative matrices

$$\mathbf{R} = \mathbf{U}_i \mathbf{R} \mathbf{U}_j, \tag{36.12}$$

where  $\mathbf{U}_i, \mathbf{U}_j$  are orthogonal matrices corresponding to any of the 24 possible symmetry operations of the point group 432. Suppose  $\mathbf{R}$  is known for one representation of a coincidence lattice with reciprocal density  $\Sigma$ . Then it follows that  $\mathbf{R}$  may be written in the form

$$\mathbf{R} = \frac{1}{\Sigma} \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix}, \tag{36.13}$$

where the matrix elements  $u_i, v_i, w_i$  are all integers and have no common factor with  $\Sigma$ . The vectors  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  give the directions into which the base vectors of the first lattice are rotated by  $\mathbf{R}$ . It may now be seen that  $\mathbf{R}$  represents a rotation about a unit vector  $\mathbf{l}$  through an angle  $\theta$  where

$$l_1 : l_2 : l_3 = w_2 - v_3 : u_3 - w_1 : v_1 - u_2 \tag{36.14}$$

and 
$$\cos \theta = (u_1 + v_2 + w_3 - \Sigma) / 2\Sigma. \tag{36.15}$$

Warrington and Bufalini (1971) have used eqns. (36.12)–(36.15) to give a computer print-out of the main coincidence lattices, their possible  $\mathbf{l}, \theta$  representations, and the corresponding matrices  $\mathbf{R}$ .

Table VII lists some of the orientation relations which give the highest density of coincidence sites in cubic crystals;  $\Sigma = 1$  implies  $\theta = 0^\circ$ , so that the lowest value of  $\Sigma$  is 3 and this corresponds to the simplest type of twin relation. In general an arbitrary boundary

TABLE VII. SOME COINCIDENCE SITE RELATIONS IN CUBIC CRYSTALS

Reciprocal density of coincidence sites. $\Sigma$	Minimum angle and axis of rotation.		Twin axes		Most densely packed planes of coincidence site lattice		Area per lattice point in coincidence in units of $b^2$	
	$\theta^\circ$	1			b.c.c.	f.c.c.	b.c.c.	f.c.c.
3	60.0	$\langle 111 \rangle$	$\langle 111 \rangle$	$\langle 112 \rangle$	{112}	{111}	1.6	0.87
5	36.9	$\langle 100 \rangle$	$\langle 012 \rangle$	$\langle 013 \rangle$	{013}	{012}	2.1	2.25
7	38.2	$\langle 111 \rangle$	$\langle 123 \rangle$		{123}	{135}	2.5	2.95
9	38.9	$\langle 110 \rangle$	$\langle 122 \rangle$	$\langle 114 \rangle$	{114}	{115}	2.9	2.6
						{111} (3) <sup>†</sup>		2.6
11	50.5	$\langle 110 \rangle$	$\langle 113 \rangle$	$\langle 233 \rangle$	{233}	{113}	3.1	1.65
13a	22.6	$\langle 100 \rangle$	$\langle 023 \rangle$	$\langle 015 \rangle$	{015}	{023}	3.4	3.6
13b	27.8	$\langle 110 \rangle$	$\langle 134 \rangle$		{134}	{139}	3.4	4.8
15	48.2	$\langle 210 \rangle$	$\langle 125 \rangle$		{125}	{157}	3.55	4.3
						{111} (5) <sup>†</sup>		4.3
17a	28.1	$\langle 100 \rangle$	$\langle 014 \rangle$	$\langle 035 \rangle$	{035}	{014}	3.9	4.1
17b	61.9	$\langle 221 \rangle$	$\langle 223 \rangle$	$\langle 334 \rangle$	{334}	{155}	3.9	3.6
19a	26.5	$\langle 110 \rangle$	$\langle 133 \rangle$	$\langle 116 \rangle$	{116}	{133}	4.1	2.2
19b	46.8	$\langle 111 \rangle$	$\langle 235 \rangle$		{235}	{235}	4.1	6.15
21a	21.8	$\langle 111 \rangle$	$\langle 124 \rangle$		{124}	{124}	6.1	4.6
21b	44.4	$\langle 211 \rangle$	$\langle 145 \rangle$		{145}	{111} (7) <sup>†</sup>	4.3	6.05

<sup>†</sup> These figures give the reciprocal density of lattice sites which are in coincidence in the close-packed planes. In all other cases, all sites in the close-packed planes are in coincidence.

will have the same ratio of normal sites to coincidence sites as the ratio  $\Sigma$  for the whole lattice, but there are also some special orientations of the boundary, parallel to densely packed planes of the coincidence lattice, which have a higher proportion of coincidence sites. Thus it is possible for all the lattice sites of a given plane to be coincidences, and there are then  $\Sigma - 1$  parallel planes which contain no coincidences. Symmetrical tilt boundaries with rational indices are always of this type, and as noted on p. 337, they may be considered to be fully coherent twin boundaries. The corresponding orientation relation may alternatively be regarded as a special case of eqns. (36.10) and (36.11) for which  $x = 0$ ,  $y = 1$ ,  $\theta = 180^\circ$ , and the rotation axis is normal to the plane of exact coincidence. When the indices of this plane satisfy the condition

$$h_i h_i = \Sigma' \quad (\Sigma' \text{ odd}) \quad (36.16)$$

the number of lattice planes traversed by a primitive lattice vector normal to the planes is either  $\Sigma'$  or  $2\Sigma'$  (see Tables II and III, pp. 38-9) but in both cases there are  $\Sigma' - 1$  planes of no coincidences in every group of  $\Sigma'$  planes so that  $\Sigma = \Sigma'$ . Alternatively, for a net with normal  $k$  rotated by  $180^\circ$ , where

$$k_i k_i = \Sigma' \quad (\Sigma' \text{ even}), \quad (36.17)$$

the number of lattice planes traversed by the primitive lattice vector parallel to  $k$  is always  $\Sigma'$ , and the value of  $\Sigma$  is then  $\Sigma = \frac{1}{2}\Sigma'$ , where  $\Sigma$  is necessarily odd as  $k_i$  are co-prime.

Equations (36.16) and (36.17) show that in general a coincidence site lattice with fixed  $\Sigma$  may be represented by rotations of  $\pi$  about two different axes but for some values of  $\Sigma$  (7, 15, 23, . . .) satisfying  $\Sigma = 4^p(8n+7)$ , where  $p$  and  $n$  are integers, there is no solution to (36.16) and there is then only one type of twin axis. Another factor is that either (36.16) or (36.17) may have more than one solution, and these solutions give rise to independent coincidence lattices which are denoted by adding  $a, b, \dots$ , to the value of  $\Sigma$ . If one of the equations has two solutions and the other has only one solution, one of the two orientation relations will have two representations as a  $180^\circ$  rotation (e.g. 13a) and the other only one (e.g. 13b).

Although a twin orientation always implies that there is a coincidence site lattice, the reverse statement is not valid. Fortes (1972) showed that a coincidence site lattice in a cubic system generated by Ranganathan's procedure will be equivalent to a  $180^\circ$  rotation about some other axis if either (a) the generating axis is parallel to a plane of type  $\{100\}$  or  $\{110\}$  or (b) the ratio  $x/y$  in eqn. (36.11) is equal to any of the relatively prime indices of the rotation axis or to the sum of all three, allowing independent choices of sign for each index. The lowest index direction which does not satisfy (a) is  $\langle 123 \rangle$ , and for rotations about this axis conditions (b) show that the smallest value of  $\Sigma$  which does not represent a twin relation is  $\Sigma = 39$ .

Planes in which every site may be a coincidence site may be expected to represent boundaries of low energy, and the most densely packed planes of this type are usually, but not invariably, one of the twin planes for the given orientation relation. However, the closest packed planes of the coincidence lattice are often relatively high index planes of the crystal lattice, so that the absolute density of lattice points is small and atoms on several of the closely spaced neighbouring parallel planes are also in the interface region. When comparing the energies of high-angle boundaries with different crystal orientations, the most important criterion is likely to be a high absolute density of coincidence sites, or equivalently a small size for the minimum period at which the boundary structure is repeated. We have therefore also included in the table the size of the minimum repeat unit for the f.c.c. and b.c.c. lattices in terms of the magnitude of the shortest lattice vector  $\mathbf{b}$ ; this corresponds, of course, to the nearest neighbour atom separation  $r_1$  in the case of the f.c.c. and b.c.c. structures.

The multiple descriptions of the orientation relation given by eqn. (36.12) lead to a corresponding multiplicity of values for the net Burgers vector in the interface given by eqns. (36.7) or (36.8). For any description, the magnitude of the Burgers vector  $|\mathbf{b}|$  has maximum and minimum values given by

$$|\mathbf{b}^{\max}| = 2|\mathbf{p}| \sin \frac{1}{2}\theta \tag{36.18}$$

and  $|\mathbf{b}^{\min}| = 2|\mathbf{p}| \cos \varphi \sin \frac{1}{2}\theta, \tag{36.19}$

where  $\varphi$  is the angle between the interface normal and the rotation axis  $\mathbf{l}$ . If the assumption is made that the description which is physically most significant is that for which  $|\mathbf{b}^{\max}|$  is smallest, the choice of the appropriate rotation matrix is straightforward. The angle  $\theta_U$  corresponding to any representation  $\mathbf{R}_U$  of eqn. (36.12) is readily obtained from the trace of the matrix which is

$$(\mathbf{R}_U)_{ii} = 1 + 2 \cos \theta_U$$

and the minimum value of  $\theta_V$  substituted into eqns. (36.18) and (36.19) then gives the dislocation density of the interface.

The criterion of minimum  $|\mathbf{b}^{\max}|$  is, not unexpectedly, equivalent to choosing the axis-angle representation of minimum rotation angle. However, the lattice relations need not be expressed as a simple rotation and if shears are also permitted the matrices  $\mathbf{U}$  of eqn. (36.12) are any unimodular matrices with integral elements. As will be shown in Section 38, this permits descriptions in which  $|\mathbf{b}|$  is further reduced, and in particular any coherent twin boundary in which the lattices are related by a simple shear may be assigned zero dislocation density. This is not a physically very useful representation when the value of  $\Sigma$  is large, so that the lattice planes parallel to the interface are sparsely populated.

Figure 8.11, which is due originally to Frank, shows the atom positions in the common  $\{111\}$  planes of two f.c.c. crystals which are rotated by  $38^\circ$  about the  $\langle 111 \rangle$  direction to produce a  $\Sigma = 7$  coincidence site lattice. The atoms are assumed to remain in coincidence site relation, and the figure illustrates the atomic configuration for two tilt boundaries of

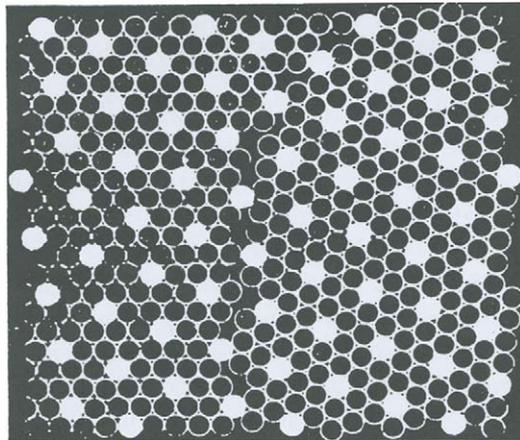


FIG. 8.11. Frank's model of a grain boundary between two f.c.c. crystals rotated  $38^\circ$  about a common  $\langle 111 \rangle$  axis (after Aust and Rutter, 1959). (Blocked-in circles represent "coincidence sites".)

different orientations. In the hard sphere model shown, it is impossible for two atoms to approach closer than the normal nearest neighbour separation, and as a result there is a considerable amount of empty volume in the vicinity of the interface. In some other models, atoms are inserted into holes of radii greater than  $0.9r_1$ , or equivalently an atom is removed only when two atoms "overlap" by more than a critical amount. However, it is clear that a still lower energy may be achieved by translation away from the coincidence site position, and the structure will then depend on the details of the interatomic forces. Atomistic calculations of this type have been made for a few boundaries, and are expected to become more important. Before discussing them, however, we take up the problem of describing the structure of a boundary which deviates by a relatively small amount from an ideal coincidence site relationship.

We find it convenient to discuss separately two kinds of deviation. The first occurs when the relations between the two lattices are such that a good fit, low-energy boundary might exist but the actual boundary is rotated slightly from this ideal orientation. The real interface can then often be considered to consist of stepped planar sections of ideal low-energy interface. The steps are discontinuities in this interface, and for symmetrical tilt boundaries they may be described as twinning dislocations. This is a simple extension of the model described in Section 32 to the generalized coherent twin boundaries introduced above. It is clear that the equivalent Burgers vectors will be smaller than a lattice vector.

The other type of deviation occurs when the orientation relation between two lattices is close to but different from a coincidence lattice relationship. The interface dislocations in the ordinary model of a low-angle boundary may be regarded as discontinuities which compensate for the discrepancy between the ideal orientation relation ( $\theta = 0^\circ$ ,  $\Sigma = 1$ ) and the actual orientation relation; we call these "primary" interface dislocations and note that their Burgers vectors must be repeat vectors of the reference lattice. In an analogous fashion, a high-angle grain boundary of good fit may be treated as an entity, and the boundary structure may be modelled by introducing "secondary" interface dislocations which accommodate the deviation from the ideal orientation relations. This concept was first formulated by Read and Shockley (1950) who used a description in terms of two arrays of dislocations. The dense array gives the mathematical description of a coincidence site, double tilt boundary in terms of a uniformly spaced set of primary interface dislocations, whilst the deviations are represented by a low-density dislocation array.

The secondary dislocation description was developed in terms of the coincidence site model by Brandon *et al.* (1964) and Brandon (1966) who linked it to field-ion microscope observations of grain boundary structures, and by Bishop and Chalmers (1968, 1971), who emphasized that boundary coincidences can be maintained even when three-dimensional coincidences vanish. The most general mathematical theory has been developed by Bollmann (1967, 1970), who has introduced a concept known as the *O*-lattice. Bollmann proposed that regions of good fit in a boundary should not be identified solely by coincidences of points of the two lattices, but also by coincidences of interior cell points which are defined so that if for any cell of one lattice the interior coordinates of a point (expressed as fractions of the cell edges  $\xi$ , as in eqn. (5.8)) are identical with the interior coordinates of that same point measured relative to a cell of the other lattice, the point is considered to be a coincidence. The *O*-lattice is the totality of such coincidences, and, unlike the coincidence site lattice, it changes continuously as the orientation is varied away from a coincidence lattice relation.

The *O*-lattice is used to discuss the primary and secondary dislocation content of any boundary, and is especially powerful in the case of interphase as distinct from grain boundaries. We thus find it convenient to defer a fuller discussion to Section 38. However, we may usefully summarize the type of dislocation model required for various grain boundary situations of interest. The simplest case to consider is when the axis of misorientation corresponds exactly to that required for a coincidence site relation, but the angular misorientation differs slightly from an exact coincidence value. The structure can then be described as that of a coincidence boundary with a superimposed sub-boundary network of dislocations

corresponding to a low angle rotation about this axis. Thus for a tilt boundary of this type, the secondary dislocations are pure edges parallel to the tilt axis, but as shown below, their Burgers vectors do not correspond to lattice vectors of either real lattice.

Deviations in the axis of misorientation  $\mathbf{r}$ , rather than in  $\theta$ , may be accommodated by a superimposed array in which the sub-boundary misorientation axis is perpendicular to the coincidence site misorientation axis. This means, for example, that a small change from a coincidence site tilt boundary is now accommodated by a secondary dislocation twist boundary. In general, deviations may involve changes in both  $\mathbf{r}$  and  $\theta$ , so that the axis of misorientation of the secondary array will lie at an arbitrary angle to the axis of the reference coincidence site relation.

Even under the most favourable conditions, these dislocation models of grain boundaries can only be regarded as approximations, and there is a major difficulty in the choice of the reference situation for secondary dislocations which will be further discussed in Section 38. Some attempts have been made to calculate the actual atomic positions in the boundary region using the interatomic force laws discussed in Section 16. There are, however, some special difficulties in making such calculations for grain boundaries.

Consider a planar grain boundary of unit normal  $\mathbf{v}$  so that the positions of the atoms in the two crystals in the regions remote from the interface are known, apart from a possible small rigid translation  $\mathbf{t}$  (see eqn. (9.1)). The interface structure may be calculated in principle by placing the atoms in a zone around the interface in arbitrary positions which are then relaxed by an iterative displacement of each atom in turn until an energy minimum is attained. The size of the crystal block used for the computer calculation must be such that periodic boundary conditions can be applied parallel to the interface, i.e. the structure of the interface must be doubly periodic in two independent directions, and the area used for the calculation must include at least one complete unit of pattern. This restricts calculations to those interfaces in which there is a reasonably sized two-dimensional net which is a section of a coincidence site lattice (ignoring any translation  $\mathbf{t}$  which may move the sites out of coincidence but does not change the periodicity). Even then, it is not obvious that a single pattern repeat unit is adequate; the results will appear more convincing if a larger area is used, and the shorter periodicity should then be revealed directly in the computed atomic positions.

The fixed positions of the atoms at large distances from the interface may be used as a boundary condition for the relaxation procedure, but themselves have to be determined because of the translation  $\mathbf{t}$ . Thus if one lattice is regarded as fixed, the atoms in the other lattice remote from the interface may be in positions corresponding exactly to a coincidence site relation or they may be translated away from these positions by any vector which joins the corner of a unit cell to an interior point of the cell. Weins (1972) has described a procedure in which the energy of the crystal is first minimized with respect to this translation, keeping all the atoms on normal sites, and then the iterative relaxation of the atoms in the boundary zone is used to find the lower energy configuration. Presumably in a complete calculation it would be necessary to iterate the displacement and atomic relaxation stages more than once, but in Weins's results the relaxations from normal lattice sites were found to be very small.

The necessity for the two-stage calculation arises from an assumed boundary condition in which the atoms remote from the interface are regarded as fixed. A possible alternative procedure is to fix the atoms at infinity on one side of the interface only and to carry out the atomic relaxation plane by plane. Any components of displacement which are common to all the atoms in a plane are immediately applied to all the succeeding planes, so that the translation at infinity is computed as the sum of the individual translations near the interface. Vitek (1970) used this method to calculate the structure of a coherent twin boundary, but its application to a general boundary is obviously more difficult.

Unlike most defect problems, it is not known at the beginning of the calculation how many atoms should be placed in the region around the interface to be relaxed. Dahl *et al.* (1972) used rigid boundary conditions and rejected overlapping atoms, i.e. those atoms with geometrical centres closer than the nearest neighbour distance, whilst Weins used a similar but opposite procedure in which the initial structure contains voids, and atoms are inserted into any holes with radii greater than 90% of the nearest neighbour distance. Hasson *et al.* (1972) tackle this problem systematically by starting with a number of atoms which is certainly too high, relaxing completely to minimum energy, and then removing the atom which has the highest energy. The structure is then relaxed again to produce a still lower energy, and the process is repeated until removal of an atom causes the energy to rise. Once again, in a full calculation it would seem necessary to iterate this process with any separate calculation of the translation vector.

There are obviously many variables in these calculations, and the best procedures have probably still to be determined; in particular it cannot be certain that the calculations do not give structures which are metastable rather than stable. However, some of the results appear to confirm the dislocation models of low-angle boundaries and the significance of periodic structures in some high-angle boundaries. They also show that the actual coincidence of boundary sites is not the important criterion in periodic boundaries since in general the lowest energy configuration corresponds to a relative translation away from the coincidence site position.

The simplest type of boundary is the fully coherent twin boundary, and one calculation for a  $\{112\}$  boundary in a b.c.c. structure has been made by Vitek (1970) using empirical, Johnson-type potentials. The work arose from the discovery (see Chapter 5) that wide monolayer twins, i.e. intrinsic stacking faults, on  $\{112\}$  planes of b.c.c. structures are mechanically unstable. Vitek therefore investigated multilayer faults, and found that a three-layer fault is metastable, but the successive displacements of the layers are not all  $(a/6)\langle\bar{1}\bar{1}1\rangle$ , as expected, but are approximately  $(a/12)\langle\bar{1}\bar{1}1\rangle$ ,  $(a/6)\langle\bar{1}\bar{1}1\rangle$  and  $(a/12)\langle\bar{1}\bar{1}1\rangle$ . The same situation was found to characterize the boundary of the infinite twin where the first atomic layer may be regarded as shifted by  $(a/12)\langle\bar{1}\bar{1}1\rangle$  with respect to the matrix and the remaining layers are each shifted by  $(a/6)\langle\bar{1}\bar{1}1\rangle$  with respect to each other. The calculated structure, Fig. 8.12, is dislocation free as expected, but there is no plane of atoms in coincidence site positions common to the two lattices. The computed structure may be derived from the coincidence site structure by a relative translation of the lattices through  $(a/12)\langle\bar{1}\bar{1}1\rangle$ ; it is also rather symmetrical and the atomic structure appears identical when viewed from the two sides of the interface. The calculations also show that the reason for the stability

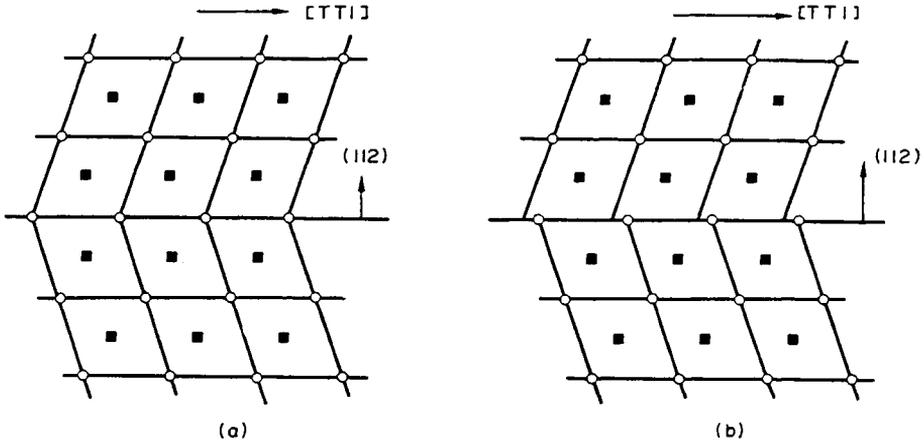


FIG. 8.12. Possible atomic structures of a {112} twin boundary in a b.c.c. material. (a) Coincidence site boundary. (b) Boundary calculated by Vitek (1970). Atoms represented by circles and squares lie in successive  $\{1\bar{1}0\}$  planes parallel to the plane of projection (after Christian and Crocker, 1980).

of the structure of Fig. 8.12(b) lies in the strong repulsive interaction.

Calculations of the structures and energies of various tilt boundaries have been described by Hasson *et al.* (1972), Weins (1972), Dahl *et al.* (1972), Smith *et al.* (1977), Pond and Vitek (1977), Pond *et al.* (1979), Crocker and Faridi (1980) and Sutton and Vitek (1980). The early workers used Morse or Lennard-Jones potentials, but Vitek and his coworkers have used a pseudopotential for aluminium and Crocker an empirical potential representing copper. Structures have been calculated for low-angle boundaries, symmetrical  $[001]$  and  $[011]$  tilt boundaries, asymmetric tilt boundaries and high  $\Sigma$  boundaries.

Figure 8.13(a) shows the computed structure for a low angle ( $10^\circ 27'$ ) symmetrical tilt about  $[100]$ , corresponding to a boundary plane  $(0, 1, 11)$  and a  $\Sigma$  value of 133. The relation to the dislocation model is obvious, and the regions of severe misfit extend only for one or two atomic distances. A higher angle ( $36^\circ 52'$ )  $(013)$ ,  $\Sigma = 5$  tilt boundary is shown in Fig. 8.13(b), and here the boundary misfit is almost continuous. Rotation of the boundary of Fig. 8.13(b) to an asymmetrical position gives the configuration shown in Fig. 8.13(c) in which the structure is much less regular, although dislocation-like groupings of atoms can still be recognized.

Most workers find an appreciable translation  $t$  away from the coincidence position, and in certain cases (Pond and Smith, 1974; Pumphrey *et al.*, 1976; Pond and Vitek, 1977) contrast effects in the electron microscope have given experimental confirmation. Ashby *et al.* (1978) and Pond *et al.* (1979) have suggested that high angle boundary structures preserve interatomic distances by the formation of randomly close-packed groups of atoms as postulated by Bernal for liquids (see p. 164).

The smallest vector representation of  $t$  may have a component normal to the interface and there are up to  $\Sigma$  positions of the boundary with different structures. Pond (1977) showed that in considering interface structure, it is preferable first to fix the boundary plane and then to reduce  $t$  to an in-plane translation  $t_i$ . There are up to  $\Sigma$  values of  $t_i$

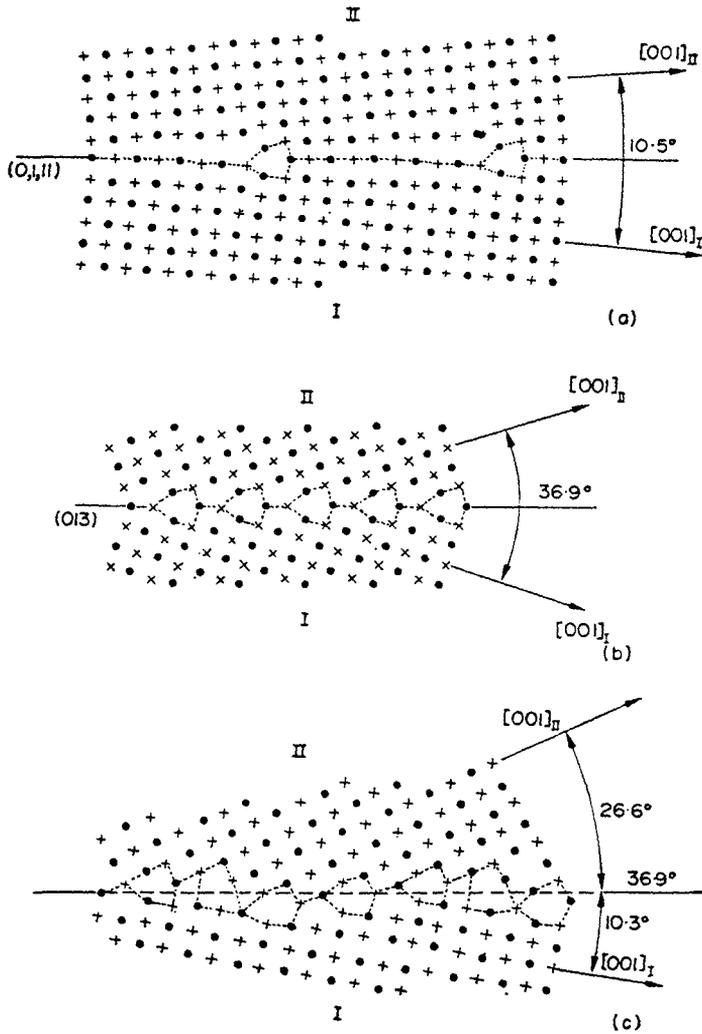


FIG. 8.13. Simulated atomic structures of some [100] tilt boundaries in aluminium. (a) Symmetric (0, 1, 11) interface,  $\Sigma = 133$ ,  $\theta = 10.5^\circ$  (b) Symmetric (013) interface,  $\Sigma = 5$ ,  $\theta = 36.9^\circ$  (c) Asymmetric  $\Sigma = 5$  interface rotated  $8.1^\circ$  from the symmetric configuration.

Atoms represented by circles and crosses lie in successive (100) planes parallel to the plane of projection (after Hasson *et al.*, 1972).

within the Wigner–Seitz cell of the two-dimensional lattice net parallel to the interface. Different in-plane translations may be symmetry related, with the same energy, or may represent structures of different energies. In either case, the junction of two such regions may be regarded as a partial grain boundary dislocation (see below).

Atomistic calculations for twist boundaries have been made by Weins *et al.* (1972), Chaudhari and Charbneau (1972), Lodge and Fletcher (1975), Guyot and Simon (1976), Bristowe and Crocker (1978) and Ingle and Crocker (1980). Crocker and coworkers found that there are three distinct symmetric structures for [001] boundaries, two of which involve in-plane translations, whereas for two families of [011] boundaries, four highly symmetric structures, three of which require in-plane translations, may arise. However, low symmetry [011] boundaries had marginally the lowest energies.

### 37. GRAIN BOUNDARY ENERGIES

The use of dislocation models enables us to calculate grain boundary energies from elastic theory; the use of this method is due mainly to Read and Shockley (1950). Consider the symmetrical, low-angle tilt boundary. The stress field of each dislocation extends for an effective distance  $Y$ , so that the energy per unit length of each dislocation will be approximately

$${}^LW_d = (B_e b/2) \ln(Y/b) + {}^LW_c, \quad (37.1)$$

where we have replaced  $r_e/r_i$  in eqn. (30.13) by  $Y/b$ . Since  $Y = b/\theta$ , and there are  $1/Y$  dislocation lines per unit length of boundary perpendicular to the lines, we have

$$\sigma = \frac{1}{2} B_e \theta \ln(1/\theta) + (\theta/b) {}^LW_c \quad (37.2)$$

as the energy per unit area of the grain boundary. This equation is usually written in the form

$$\sigma = \sigma_0 \theta (A - \ln \theta), \quad (37.3)$$

where  $\sigma_0$  and  $A$  do not depend on  $\theta$ . The parameter  $\sigma_0$  may be calculated from elastic theory, but  $A$  contains the unknown core energy of the dislocations.

This derivation is intuitive, and we have given no formal justification for the assumption that the extent of the stress field is about equal to  $Y$ . In their original paper, Read and Shockley summed the stress field of all the dislocations to find the shear stress acting on the slip plane of a particular dislocation in the array. They thus found the energy of this dislocation from the work required to create it by a virtual process. We shall follow a later, simpler method, which may be illustrated by reference to Fig. 8.14.

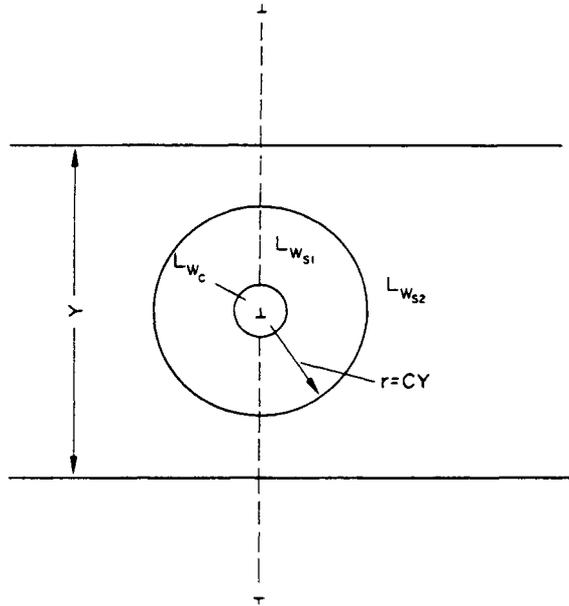


FIG. 8.14. To illustrate the calculation of grain boundary energy (after Read and Shockley, 1950).

We divide the crystal into uniform parallel strips, each containing one dislocation and having energy  ${}^LW_d$ . In calculating  ${}^LW_d$ , we first write, as before,

$${}^LW_d = {}^LW_s + {}^LW_c,$$

where  ${}^LW_c$  is the core energy of the bad crystal at the centre of the dislocation line. We then subdivide  ${}^LW_s$  into the elastic energy in a cylindrical region around the dislocation line ( ${}^LW_{s1}$ ) and the energy in the volume of the strip outside this region ( ${}^LW_{s2}$ ). The radius of this cylindrical region is  $r = CY$ , where the constant  $C < 1$  is not precisely defined, but must be large enough to ensure  $r \gg b$  and small enough for the stress field inside the cylinder to be approximately equal to that of the enclosed dislocation alone.

Read and Shockley now consider a small change in which the boundary angle decreases by  $d\theta$ , leading to increases in  $Y$  and  $r$  given by

$$-d\theta/\theta = dY/Y = dr/r.$$

The energy  ${}^LW_c$  is obviously unaffected by the change in spacing, so the change in energy is given by the sum of  $d{}^LW_{s1}$  and  $d{}^LW_{s2}$ . Consider first  $d{}^LW_{s2}$ . The volume of this region increases, but the energy in a small unit volume decreases because the dislocations move further apart. The volumes of corresponding small cylindrical elements with axes parallel to the dislocation line vary as  $Y^2$  as  $\theta$  changes. A dimensional argument shows that the energy density in each such volume varies as  $b^2/Y^2$ , so that the strain energies of corresponding cylindrical elements are unchanged. The two effects thus exactly balance and  $d{}^LW_{s2} = 0$ .

The volume of the region of energy  ${}^LW_{s1}$  is increased as  $\theta$  decreases. Since this region is chosen to lie within the effective stress field of the single dislocation, the strain energy density is unchanged. The change in  ${}^LW_{s1}$  per unit length of dislocation line is thus equal to the strain energy in the annular region of the cylinder of unit length enclosed by radii  $r, r+dr$ . The self-energy of a dislocation is equal to the work done on the slip plane in the virtual process of creating the dislocation (p. 264); the energy in the annular region is thus equal to the work done on the part of the slip plane of area  $dr$  per unit length of line. This result follows because the work done over the spherical surfaces of the annular region vanishes.

The dislocation is created by a relative movement  $\mathbf{b}$  on the slip plane, work being done against the shear stress acting on the slip plane and in the slip direction. During the virtual process, the stress at any point rises from its initial value zero to its final value  $X_{12}$ , so that the work done on the part of the slip plane  $dr$  is  $\frac{1}{2}X_{12}b dr$ . Substituting from eqn. (30.7) for  $X_{12}$  we have<sup>†</sup>

$$d{}^LW_d = d{}^LW_{s1} = (B_e b/2r) dr = -(B_e b/2\theta) d\theta$$

$$\text{or} \quad {}^LW_d = \frac{1}{2}B_e b(A - \ln \theta), \quad (37.4)$$

where  $A$  is an undetermined constant of integration. From this, we obtain as before

$$\sigma = \frac{1}{2}B_e \theta(A - \ln \theta), \quad (37.5)$$

which is identical with (37.3).

As previously mentioned, the constant term includes the unknown core energies of the dislocations, but the form of the above equation has been deduced from elastic theory alone. The derivation shows that  $A$  is determined not only by the core energy  $W_c$ , but also by the constant parts of the elastic energy, such as that contained in  $W_{s2}$ . In eqn. (37.3), we have used the shear stress calculated from isotropic elasticity, but Read and Shockley have shown that the shear stress on the slip plane at a distance  $r$  from a straight dislocation line can always be written

$$X_{12} = \tau_0 \frac{b}{r} + \tau_1 \frac{b^2}{r^2} + \dots,$$

where  $\tau_0$  is uniquely determined by the anisotropic elastic constants. The coefficients  $\tau_1, \dots$ , are determined by the interaction of good material near the core with the bad material of the core itself, and may thus be regarded as corrections when linear elasticity theory is no longer applicable. These terms contribute negligibly to the shear stress at large  $r$ , and the extra energy which they represent may thus be included in  ${}^LW_c$  since it also is localized near the dislocation. The general expression for the energy of a small angle tilt boundary is then

$$\sigma = \frac{b\tau_0}{2} \theta(A - \ln \theta). \quad (37.6)$$

An estimate of the value of  $A$  was made by van der Merwe (1950) who developed a more rigorous theory of the grain boundary energy in which he used the Peierls-Nabarro approxi-

<sup>†</sup> Note that  $r, \theta$  as used here have different meanings from the cylindrical coordinates  $r, \theta$  of (30.8).

mation (p. 273). With typical values of the elastic constants, he obtained  $A \approx 0.5$ . Experimental results seem to fit the equation quite well for  $A \approx 0.2$ , in very good agreement with this.

Equation (37.3) shows that  $\sigma$  approaches zero as  $\theta \rightarrow 0$ , as is physically obvious. The energy per dislocation,  $b\sigma/\theta$ , however, becomes infinite, in agreement with the previous conclusion that the energy of a single dislocation in an infinitely large crystal is infinite. The derivation assumes that the spacing of dislocations is uniform, but as discussed above this is only possible at a number of discrete angles. For intermediate angles, the irregularities in the array cause extra energy terms. For a small deviation  $\delta\theta$  from the nearest rational angle  $\theta = 1/m$ , the extra energy is of order

$$-\frac{\sigma_0 \delta\theta}{m} \ln \delta\theta.$$

As  $\delta\theta \rightarrow 0$ , the slope of the  $\sigma$ - $\theta$  curve thus becomes infinite, and the rational angles  $\theta = 1/m$  represent cusps in the true curve of  $\sigma$ - $\theta$ . Equation (37.3) gives the locus of these cusps (Fig. 8.15). The point  $\theta = 0$  is one such cusp, and the slope of the energy curve is

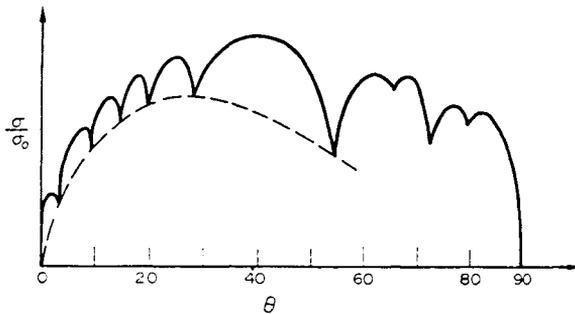


FIG. 8.15. Variation of boundary energy with orientation (after Read and Shockley, 1950). (The deep cusp at  $53^\circ$  represents a  $\{210\}$  twin boundary.)

infinite at the origin. As  $\theta \rightarrow 0$  the density of the cusps increases, so that in the low-angle region eqn. (37.3) effectively gives the true variation of energy with orientation. The depths of the low-angle cusps are correspondingly small; the only important cusps are those of low indices, which correspond to the usual coherent twin boundaries. The relative rotation of the lattice is then too large for the energy formula to apply in any case.

In the above we have discussed the energy of a tilt boundary with only one degree of freedom. If we next consider the asymmetrical tilt boundary of Fig. 8.7, we have to sum the energies of two sets of dislocations. This energy may be divided into three parts, namely an energy per dislocation  ${}^LW_{100}$  which the first array would possess in the absence of the second array, a similar energy  ${}^LW_{010}$  for the second array, and an energy per dislocation which results from the interactions of the stress field of one array with the strain field of the other. The previous method of calculation, however, shows that the interaction energy can only contribute to the terms  ${}^LW_{s2}$ . The argument above may thus be repeated separately

for each set of dislocations, and the interaction energy absorbed in the constants of integration. This gives

$$\sigma = ({}^LW_{100}/Y_{100}) + ({}^LW_{010}/Y_{010}) = \sigma_0\theta(A - \ln \theta) \quad (37.7)$$

as before. The term  $\sigma_0$  now depends explicitly on  $\varphi$ ; using eqns. (36.2) and (36.3)

$$\sigma_0 = \frac{b\tau_0}{2}(\cos \varphi + \sin \varphi), \quad (37.8)$$

and  $\sigma_0$  is proportional to the total density of dislocations. This is true only if the two sets have equal strengths  $b$  and equal values of  $\tau_0$ . The constant of integration  $A$  is now also a function of  $\varphi$ , and formulae have been given for its variation (Read and Shockley, 1950).

The argument may be further extended to the general small-angle grain boundary, in which the two lattices are related by a rotation  $\theta$  about an arbitrary axis  $\mathbf{l}$ . In the last section we described how this boundary could be described by three or more sets of dislocation lines. The spacing of each set of lines  $Y_i$  was shown to be inversely proportional to  $\theta$ ; we write  $Y_i = 1/C_i\theta$  so that  $C_i\theta$  gives the density of lines of Burgers vector  $\mathbf{b}_i$ . The energy per dislocation of this set is then

$${}^LW_i = \frac{\tau_{0(i)}b_{(i)}^2}{2}(A_{(i)} - \ln \theta), \quad (37.9)$$

where the term  $A_i$  includes the energies of interaction of the  $i$ th set with all the other dislocations. The total energy of the boundary is now

$$\sigma = C_i\theta {}^LW_i \quad (37.10)$$

which is of the form (37.3) with

$$\sigma_0 = \frac{1}{2}\tau_0 C_i b_i^2. \quad (37.11)$$

The predicted variation of  $\sigma$  with  $\theta$  is thus of the same form for all types of low-angle grain boundary, and experimental results are in very good agreement with the theoretical curve. The agreement, indeed, is generally too good, since it extends to much higher angles than would be expected. As we have emphasized, the theoretical derivations are only valid for widely spaced dislocations, but the range of the formula can be extended by supposing  $A$  to be not constant but a function of  $\theta$ . Certain qualitative features of high-angle boundaries have already been mentioned, namely the existence of deep energy cusps, corresponding to coherent twin boundaries. It should also be mentioned that different choices of the reference lattice, which lead to different dislocation descriptions of a boundary, will give different apparent energies if eqn. (36.7) is used. Thus two grain boundaries related by a relative rotation matrix  $R$  may also be described as related by  $R_1 R R_2$ , where  $R_1, R_2$  are rotations which leave the lattice in an identical orientation. Since these descriptions represent the same physical situation, the energies must actually be equal. The anomaly arises because at least one of the descriptions must be in terms of a large rotation and the values of  $A$  must thus be such as to give the same energy.

Although dislocation calculations of grain boundary energies are valid only for low-angle boundaries, they may also be applied in principle to secondary dislocation distributions, in which case they give the excess energy of the boundary over that of a reference (coincidence site) boundary. However, the difficulties in assigning specific Burgers vectors to these dislocations make this calculation very uncertain, and Li (1972) has proposed a slightly different approach based on the disclination model of Fig. 8.10. From that figure, we see that

$$\tan \frac{1}{2}\theta = b/2Y' = L\omega/Y'_\omega, \quad (37.12)$$

where  $b$ ,  $Y'$  as before are the Burgers vector and projected spacing of the dislocations, and  $\omega$ ,  $L$ ,  $Y'_\omega$  are respectively the disclination strength, the separation of the dipole elements, and the separation of the disclination dipoles. The strain energy of an isolated disclination dipole is the same as that of the equivalent dislocation wall, and may be expressed (Li, 1960) as

$${}^LW_\omega = \{\mu\omega^2L^2/4\pi(1-\nu)\} \ln(r_e/L). \quad (37.13)$$

Using a model in which infinite dislocation walls of identical spacing are added together to form a high-angle boundary in such a way that the dislocations cluster into disclinations, Li obtains an expression for the variation of energy with misorientation for situations intermediate between two low-energy configurations. The energy appears as a sum and the disclinations are approximated as walls containing  $M$  dislocations with separations of  $N$  Burgers vectors. The angle of the boundary may be varied by varying  $L$  or  $Y'_\omega$  or the two simultaneously, eqn. (37.12), and the excess energy is then zero for  $M = 1$  and  $M/N = 1$ , and reaches a maximum for  $M/N \sim \frac{1}{2}$ . Further work is needed to test the validity of this approach.

### 38. INTERPHASE BOUNDARIES: SURFACE DISLOCATIONS

The boundary between two crystals of different structure also has five degrees of freedom in the general case, and its structure and behaviour is similar in many ways to that of a grain boundary. Three degrees of freedom are needed to specify the relative orientations of the two lattices, but of course it is no longer possible to carry one lattice into the other by means of a pure rotation. As discussed in Section 9, the lattices are now related by the general linear transformation  $\mathbf{RP}$ , and may also be described by any other linear transformation  $\mathbf{U}_i\mathbf{RPU}_j$ , where  $\mathbf{U}_i$  and  $\mathbf{U}_j$  as in eqn. (36.12) represent either symmetry operations of the lattice or other lattice-preserving deformations.

In discussing the physical structure of a boundary, we shall find it convenient to distinguish three types of interface which we describe as incoherent, semi-coherent, and coherent respectively. The incoherent phase interface is the analogue of the high-angle grain boundary. There are no continuity conditions for lattice vectors or planes to be satisfied across the interface, the structure of which is relatively disordered. Although a formal description of this boundary may be given in terms of dislocation theory, this has no physical significance, and there is no lattice correspondence when the boundary moves.

The fully coherent interface is to be compared with the meeting of two twins along their composition ( $K_1$ ) plane. The lattices match exactly at the interface, and "corresponding"

lattice planes and directions are continuous across the interface, although they change direction as they pass from one structure to another. We emphasize here that a coherent interface in the sense in which we use the term does not necessarily imply a rational interface; in mechanical twinning, for example, the coherent interface may be rational or irrational, depending on the type of twinning. As we have seen in Section 9, it is not generally possible to find a coherent interface between two arbitrary structures. A necessary condition for a plane of exact matching to exist is that the deformation relating selected unit cells of the two structures should have a zero value of one principal strain, and since this implies a relation between the cell parameters it can only be satisfied coincidentally. The best-known examples of nearly exact matching are those involving f.c.c. and h.c.p. structures with virtually identical interatomic distances in the octahedral and basal planes respectively. If the matching condition is nearly satisfied, two phases may be forced elastically into coherence across a planar boundary. This is usually possible only when one (included) crystal is very small, since the stress at any point increases with the size of the crystal. Forced elastic coherence of this kind may exist at the nucleation or early growth stage of a transformation, the elastic strains near the interface being possibly much larger than the normal elastic limit.

We now turn to the semi-coherent interface, the prototype of which is the low-angle grain boundary. Such an interface consists of regions in which the two structures may be regarded as being in forced elastic coherence, separated by regions of misfit. A fully coherent interface is possible only when the deformations generating the two lattices from some reference lattice are compatible in ordinary space, in the sense discussed in Section 34. When these deformations are not compatible, we must specify an additional (lattice invariant) deformation to enable us to fit the regions together in ordinary space. If we use the theory of the continuous distribution of dislocations, we can do this for any two lattices, but in many cases the resultant surface density of dislocations will be extremely high. In cases where the density is low enough for individual dislocations in the interface to accomplish the lattice matching, the description acquires a physical significance, and we may have a semi-coherent boundary. When such a boundary exists, we cannot specify an overall correspondence (lattice generating deformation) relating one lattice to the other, but a local correspondence of this kind exists. This is exactly the analogy of the relation between the reference lattice and the dislocated lattice discussed on p. 317.

Note that in the sense used here, a semi-coherent boundary is a particular type of plane interface. A curved boundary, as in lenticular plates, necessarily contains additional misfit regions. If the curvature is small, these may be considered as steps of atomic height, and correspond to the twinning dislocations and transformation dislocations already discussed in Section 32. These discontinuities, which can glide in the composition or habit plane, are quite distinct from the discontinuities needed to correct the mismatch of the two lattices.

During a phase transition, one crystal grows at the expense of the other by the migration of the boundary interface. It is clear that if the structures are fully coherent a lattice correspondence is implicit in this process, so that labelled rows or planes of lattice sites in one crystal become rows or planes in the other. We write lattice sites rather than atoms since only lattices and not structures are related by affine transformations; in the simplest cases,

there will be a correspondence of atomic sites. The motion of a coherent boundary thus produces a macroscopic change of shape which is specified by the relation between the lattices.

If the boundary is incoherent, there is no correspondence and no shape change when it moves. Thus the growth of a mechanical twin changes the shape of the specimen, but the motion of a grain boundary does not. A semi-coherent boundary requires more detailed discussion. Since the lattice deformations are not compatible, the motion of the boundary cannot produce a change of shape given by the local lattice correspondence. In fact, the shape change produced is obtained by combining the lattice change with the change due to the migration of the discontinuities in the boundary (lattice invariant deformation) exactly as on p. 322. It follows that for both fully coherent and semi-coherent boundaries the shape change produced by motion of the boundary is an invariant plane strain.

We shall now discuss a simple model of a semi-coherent boundary, before giving a more general treatment of the structure of such a boundary. Suppose first that the boundary contains a single set of parallel dislocations with a common Burgers vector. If this vector does not lie in the interface, or if the dislocations are pure screws, we may suppose that the corresponding glide planes for the dislocations in the two structures meet edge to edge in the interface. The line of intersection, which is parallel to the dislocations, is an invariant line of the lattice deformation (see p. 63). With a single set of parallel dislocations, the lattice invariant deformation produced by motion of the dislocation array is a simple shear, and the boundary we have just described is the usual dislocation model of a martensitic type interface. If the dislocations are pure edges, the lattice deformation reduces to a pure rotation, and the interface is a symmetrical low-angle tilt boundary, as already discussed.

If the direction of the Burgers vector is in the interface plane and is not parallel to the dislocation lines, we have a different type of boundary. The habit plane is now the glide plane, so that as the boundary is displaced the dislocations in it must climb. The motion is thus non-conservative in the sense that if the boundary sweeps through a region of one crystal, the number of atoms incorporated into the other crystal will differ from the number originally present in the swept region. The lattices again match in the direction of the dislocation lines, but the density of equivalent lattice points<sup>†</sup> in planes parallel to the habit plane is different in the two structures. The energy of a boundary of this type was calculated by van der Merwe (1950) and Brooks (1952); it may be called "epitaxial" since the structure envisaged is similar to that postulated for epitaxial layers deposited on crystals.

Figure 8.16 shows a simple epitaxial semi-coherent boundary in which the dislocations are of edge type. As the boundary moves towards the top of the figure, the region  $ABCD$  of the upper crystal is transformed into a region of lower crystal with shape  $D'C'BA$ . This macroscopic shape change is an invariant plane strain, and results from a combination of the shape change of a unit cell with the uniaxial contraction produced by motion of the dislocation array. However, we may note an important difference between the motion of this type of semi-coherent boundary and that of boundaries of the martensitic type. The

<sup>†</sup> We define equivalent lattice points as lattice points outlining unit cells in the two structures containing the same number of atoms.

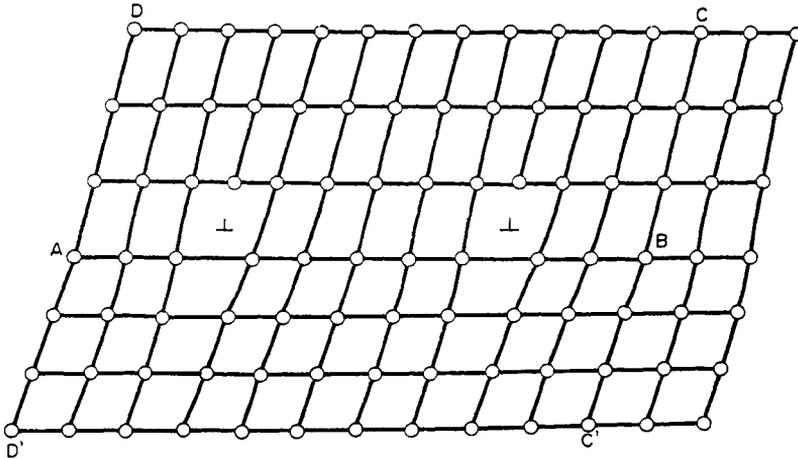


FIG. 8.16. To illustrate a semi-coherent interphase boundary containing discontinuities of edge-dislocation type.

dislocation climb associated with the motion of the boundary in Fig. 8.16 requires atomic displacements through distances of the order of the distance moved by the boundary. The motion will thus be relatively slow, and may be accompanied by a transfer of atoms from one side of the growing crystal to the other, thus effectively destroying the correspondence and its associated shape change. In the particular case of Fig. 8.16 the shape change could be minimized by a net flow of atoms from right to left as the boundary moves upwards, sites being destroyed on the right of the growing crystal and created on the left. The extra strain energy associated with the shape change within a constraining matrix should provide a driving force for this motion (Christian, 1962). Thus if epitaxial semi-coherent boundaries exist, it is rather improbable that their motion will produce observable shape changes in the same way as the motion of martensitic semi-coherent boundaries.

We have considered only a single set of parallel dislocations, but certain other boundaries may be envisaged in which the dislocation array is equivalent to a single set of dislocations. These either contain parallel sets of dislocations with differing Burgers vectors, or sets of dislocations in different directions but with a common Burgers vector. If the individual sets of dislocations are all able to glide in planes of the two lattices which intersect the habit plane, the interface will be of the martensitic type. The lattice invariant deformation produced by the glide motion of the dislocations is then again a simple shear, although the shear direction (in the first case) or the invariant plane of the shear (in the second case) will generally be irrational in both lattices.

Dislocation arrays which reduce to a single array with a Burgers vector in the interface again represent an epitaxial boundary with a single direction of lattice fit. A more general boundary of this type is obtained by combining two or more sets of non-parallel interface dislocations, each set having a Burgers vector in the interface. The whole dislocation content of the boundary may then be replaced by a crossed grid of edge dislocations, compensating for the different densities of the equivalent lattice points in two mutually perpendicular

directions. This is a simple extension of the boundary shown in Fig. 8.16, and van der Merwe showed that its energy is similar to that of a low-angle grain boundary. The shape deformation is again an invariant plane strain, but the habit plane no longer contains an invariant line of the lattice deformation.

In the above discussion, we have emphasized the distinction between a boundary in which the dislocations can glide and one in which they must climb as the boundary moves. More generally, we may distinguish between boundaries which are dependent on thermally activated processes for their motion, and boundaries which are mobile under suitable driving forces even at very low temperatures. The latter type of boundary may be called "glissile". A necessary condition for a boundary to be glissile is that its motion be conservative.

We now turn to the general dislocation description of the boundary between any two lattices in given orientation. As on p. 342, we may specify a net dislocation content, although its resolution into individual dislocation lines is not necessarily unique. The dislocation content may be found from first principles, or by utilizing the theory of continuous distributions of dislocations.

Consider first the argument we used on p. 342, We may repeat this exactly for the case in which the two real crystals have different structures if we replace  $\mathbf{R}_+$  by  $\mathbf{S}_+$ , the deformation carrying the reference lattice into the lattice of one crystal in its final orientation, and  $\mathbf{R}_-$  by  $\mathbf{S}_-$ , with a similar interpretation. We thus have for the total Burgers vector of the dislocations crossing any unit vector  $\mathbf{p}$  in the interface

$$\mathbf{b} = (\mathbf{S}_+^{-1} - \mathbf{S}_-^{-1})\mathbf{p}. \tag{38.1}$$

In particular, if we take the reference crystal to have the structure and orientation of the (-) crystal, we have

$$\mathbf{b} = (\mathbf{S}^{-1} - \mathbf{I})\mathbf{p} = (\mathbf{P}^{-1}\mathbf{R}^{-1} - \mathbf{I})\mathbf{p}, \tag{38.2}$$

where  $\mathbf{S} = \mathbf{R}\mathbf{P}$  is the lattice deformation relating the two crystals, and  $\mathbf{P}$  and  $\mathbf{R}$  are a pure deformation and a pure rotation respectively.

Alternatively, suppose we have a continuous distribution of dislocations specified by the tensor  $A_{ij}$  (p. 316), and concentrate all the dislocations into a shell of thickness  $t$  so that  $A_{ij}$  vanishes outside this shell. Now let  $t \rightarrow 0$  and  $A_{ij} \rightarrow \infty$  in such a way that the product  $tA_{ij}$  remains finite and tends to  $B_{ij}$ . We describe  $B_{ij}$  as the surface dislocation tensor. By applying Stokes's theorem to a small circuit intersected by the dislocation shell, we find (Bilby, 1955)

$$B_{ij} = -\epsilon_{jkl}v_k(\mathbf{S}_{+il}^{-1} - \mathbf{S}_{-il}^{-1}), \tag{38.3}$$

where  $\mathbf{v}$  is the positive normal to the interface plane. It should be noted that in accordance with our usual practice in Chapter 2 and elsewhere,  $\mathbf{S}_+$  and  $\mathbf{S}_-$  in eqns. (38.1) and (38.3) (when written as matrices) transform the components of a reference lattice vector into those of a vector in the real lattices. They are thus not directly identifiable with  $\mathbf{D}$  of eqn. (34.3), which gave the relation between the lattice bases, and in terms of the notation used in Section 34 we should have  $\mathbf{S}^{-1} = \mathbf{D}'$ , i.e.  $S_{+il}^{-1} = D_{+ii}$  where  $\mathbf{D}_+$  and  $\mathbf{D}_-$  are the values of  $\mathbf{D}$  in the two crystals.

Now consider the dislocations which cross a small area of the shell defined by the vector  $\mathbf{p}$  and the vector thickness  $t\nu$ . The normal to this area has components  $dS_j = t\varepsilon_{jmn}p_m\nu_n$ , and using eqn. (34.2) and putting  $tA_{ij} = B_{ij}$  in the limit,

$$b_i = B_{ij}\varepsilon_{jmn}p_m\nu_n \quad (38.4)$$

for the resultant Burgers vector. Since we have now allowed  $t$  to tend to zero, this is the net Burgers vector of the dislocation lines which cross  $\mathbf{p}$ . Substituting from (38.3), and using the relation

$$\varepsilon_{jkl}\varepsilon_{jmn} = \delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm},$$

we obtain finally

$$b_i = (S_{+il}^{-1} - S_{-il}^{-1})p_l, \quad (38.5)$$

which is identical with (38.1).

The surface dislocation tensor  $B_{ij}$  gives the  $i$ th component of the resultant Burgers vector of the dislocation lines cutting unit length of line in the surface perpendicular to the  $j$ th direction. When the dislocation content of the surface is smoothed out (continuous), we may think of the whole array as a single defect, called by Bilby a "surface dislocation". The tensor  $B_{ij}$  is the analogue for a surface dislocation of the Burgers vector of a line dislocation.

We have emphasized that for a fully coherent boundary, the lattice deformation must be an invariant plane strain. Two lattices are related by such a strain if  $\mathbf{S}_+$  has the form of eqn. (9.10) and  $\mathbf{S}_- = \mathbf{I}$ . Substituting into (38.3) we find that  $B_{ij}$  is identically zero, proving again that no dislocations are required in the interface when the lattice deformations are compatible (Basinski and Christian, 1956; Bilby and Smith, 1958). This applies, of course, to any twin boundary, but difficulties arise in the case of symmetrical low-angle tilt boundaries which may be regarded as twin boundaries for coincidence site relations of high  $\Sigma$  (Section 36). Although the lattice points may be formally related by a simple shear, the dislocation model obtained by treating  $\mathbf{S}$  as a rotation is physically more significant than a zero-dislocation model (see p. 350). In fact it is very difficult to formulate rules for the specification of the dislocation content of any boundary unless restrictions are placed on  $\mathbf{S}$ . Figure 8.17, for example, shows the same grain boundary as Fig. 8.9, but the deformation  $\mathbf{S}$  is no longer taken to be a pure rotation. The net Burgers vector defined by eqn. (38.1) in the same reference lattice as that shown in Fig. 8.9 has been changed, although the positions of all lattice points are identical; in the notation used by Hirth and Balluffi (1973) and some others, the boundary is composed of intrinsic grain boundary dislocations (IGBDs) and its Burgers vector content cannot be made specific.

It might seem obvious that we should find the minimum value of  $|\mathbf{b}|$  for all possible descriptions of the lattice relation, just as we found it for all descriptions which are rotations in Section 36. This is closely related to a procedure suggested by Bollmann for choosing the "correct" representation  $\mathbf{S}$ , and although we shall find that this is not valid without further restrictions, it is convenient to develop further the elegant formalism of his  $O$ -lattice theory, which has already been mentioned on p. 351.

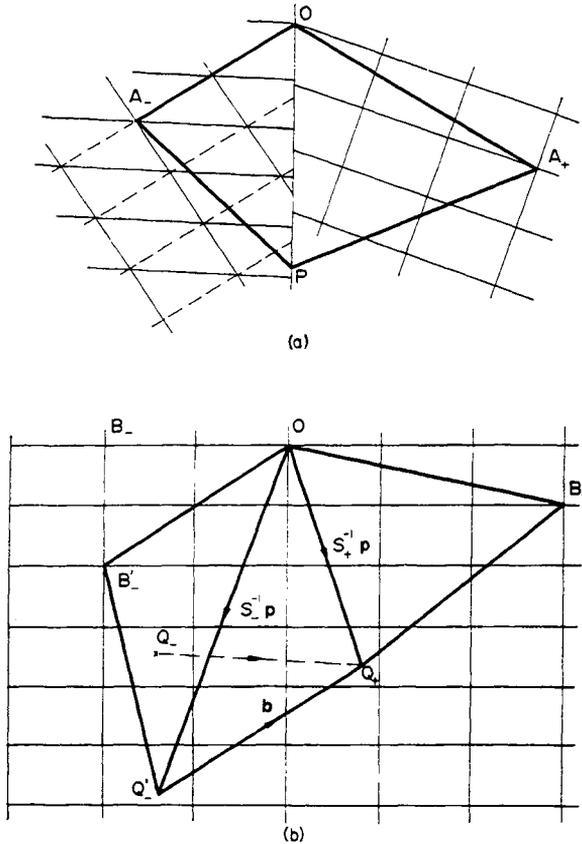


FIG. 8.17. To illustrate how a redefinition of the lattice relations in the grain boundary of Fig. 8.9 gives a different result for the net Burgers vector of the dislocations crossing  $OP$ . The + part of the circuit shown in (b) is identical with that in Fig. 8.9 but the reference unit cell is now related to a different cell of the - lattice (shown by the unbroken lines in (a)), so that  $S_-$  in (38.1) is no longer a pure rotation. The closure failure in (b) is now  $Q'_-Q_+$ ; the previous closure failure  $Q_-Q_+$  is shown by a broken line.

For two lattices related by a general linear transformation  $y = Sx$ , a point  $x^{(0)}$  of the  $O$ -lattice is defined by the simultaneous conditions

$$Sx = x + u = y = x^{(0)}, \tag{38.6}$$

where  $x = w + \xi$  is any vector and  $w$  and  $u$  are lattice vectors of the first (or -) lattice which is used as the reference lattice. The conditions ensure that the end point of the vector  $x^{(0)}$  has the same internal coordinates  $\xi_i$  in the original basis and in the "natural" basis of the second lattice which is related though  $S$  to the original basis.

Eliminating  $x$  from eqn. (38.6) gives

$$u = y - x = (I - S^{-1})x^{(0)}, \tag{38.7a}$$

which is identical with (38.2) if  $\mathbf{b}$  is replaced by  $-\mathbf{u}$  and  $\mathbf{p}$  is replaced by  $\mathbf{x}^{(0)}$ . However, the interface has not yet been introduced and the vectors  $\mathbf{x}^{(0)}$  are not restricted to a plane; in fact, the equation is regarded as a definition of these vectors  $\mathbf{x}^{(0)}$  through

$$\mathbf{x}^{(0)} = (\mathbf{I} - \mathbf{S}^{-1})^{-1} \mathbf{u}. \quad (38.7b)$$

By substituting for  $\mathbf{u}$  the three base vectors of the reference ( $-$ ) lattice in turn, we see that the columns of the matrix  $(\mathbf{I} - \mathbf{S}^{-1})^{-1}$  define the corresponding base vectors of the  $O$ -lattice. Any non-coplanar triad of vectors  $\mathbf{u}$  specifies a primitive or non-primitive cell of the reference lattice and eqn. (38.7b) then gives the corresponding  $O$ -lattice cell; the ratios of the volumes of the two cells is given by  $|\mathbf{I} - \mathbf{S}^{-1}|^{-1}$ .

The matrix  $(\mathbf{I} - \mathbf{S}^{-1})$  may be of rank 3, 2, or 1 (or trivially zero when the  $+$  and  $-$  lattices coincide). When it is of rank 3, the solutions (38.7) represent a lattice of points in three dimensions, and each point  $\mathbf{u}$  of the reference lattice has an "image" which is a point of the  $O$ -lattice. In other cases, the  $O$ -lattice becomes partly continuous. Thus if the matrix is of rank 2, solutions only exist for reference lattice points  $\mathbf{u}$  lying in a plane through the origin, and each such point has an image which is a line containing an infinite number of  $O$ -lattice points (an  $O$ -line). The  $O$ -lattice then degenerates into a set of parallel lines. The direction of these lines is one in which the two lattices fit together exactly; i.e.  $\mathbf{S}$  is an invariant line strain (p. 63). The  $O$ -lines are parallel to the invariant line and the independent components of  $\mathbf{u}$  are confined to the plane with the invariant normal. Similarly, if the rank of  $(\mathbf{I} - \mathbf{S}^{-1})$  is 1, the  $O$ -lattice degenerates into a set of parallel planes and the relation  $\mathbf{S}$  represents an invariant plane strain. It is readily seen that when  $\mathbf{S}$  has the form (9.10), each  $2 \times 2$  sub-determinant of  $(\mathbf{I} - \mathbf{S}^{-1})$  is necessarily zero and hence the rank of this matrix is 1.

As already noted, there are many possible choices of  $\mathbf{S}$  which will relate two lattices in fixed relative orientations. Bollmann postulated that the correct choice is that which maximizes the three-, two- or one-dimensional unit cell of the  $O$ -lattice; this means that the distance between corresponding points  $\mathbf{y}$  and  $\mathbf{x}$  of eqn. (38.6) is minimized, and that  $\mathbf{S}$  relates nearest neighbours in the region round the origin. (Note that whereas a coincidence site lattice for fixed real lattices is independent of the choice of  $\mathbf{S}$ , the  $O$ -lattice varies with  $\mathbf{S}$ .) Since the highest non-zero determinant contained in  $(\mathbf{I} - \mathbf{S}^{-1})$  gives the ratio of the volume, area, etc., of the unit cell of the reference lattice to that of the  $O$ -lattice, the condition is equivalent to finding the smallest absolute value of this determinant.

All rotations are invariant line strains and hence give  $O$ -line lattices. In a plane normal to the rotation axis, the rotation has a simple two-dimensional representation  $\mathbf{A}$ , and in an orthonormal coordinate system

$$(\mathbf{I} - \mathbf{A}^{-1})^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \cot \frac{1}{2}\theta \\ -\frac{1}{2} \cot \frac{1}{2}\theta & \frac{1}{2} \end{pmatrix}. \quad (38.8)$$

Figure 8.18 shows the orientations of two cubic lattices rotated about a cube edge through  $\theta = 2 \arctan \left( \frac{3}{20} \right)$  and the corresponding planar section of the  $O$ -line lattice. The figure illustrates an important geometric property of the  $O$ -lattice, namely that each  $O$ -point may act as an origin for the relation  $\mathbf{y} = \mathbf{S}\mathbf{x}$  which then relates nearest neighbour points

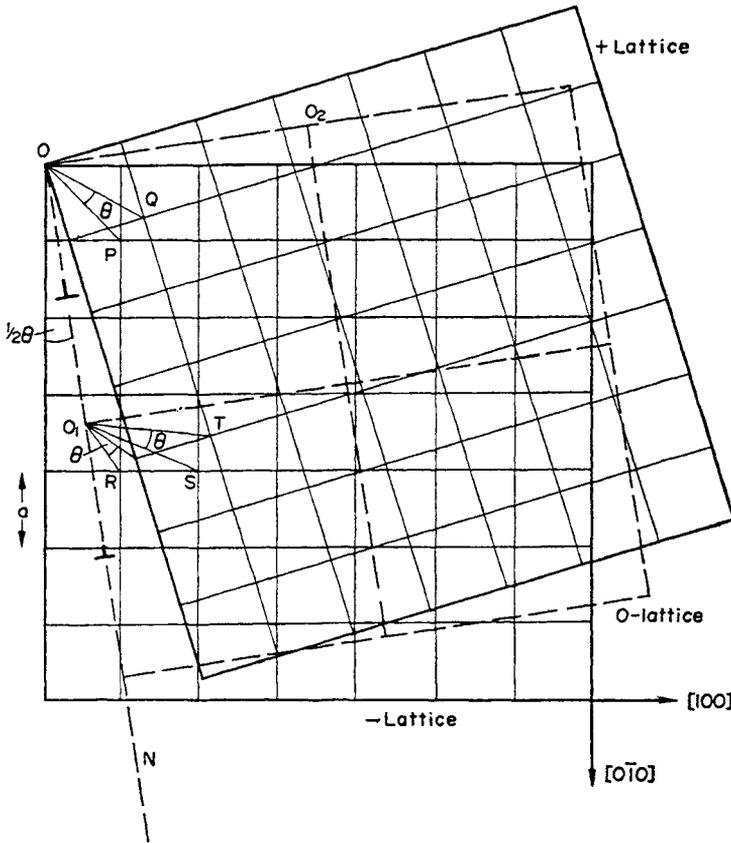


FIG. 8.18. (001) section through the  $O$ -line lattice for two cubic crystals rotated about  $[001]$  through  $\theta = 2 \arctan(\frac{3}{20})$ . All points of the  $+$  lattice may be obtained from those of the  $-$  lattice by a rotation of  $\theta$  about the origin  $O$ ; e.g. this carries  $OP$  into  $OQ$  and  $OR$  into  $OT$ . Nearest neighbour points of the two lattices are related by a rotation of  $\theta$  about the nearest point of the  $O$ -lattice; e.g.  $OP$  into  $OQ$  and  $O_1S$  into  $O_1T$ .

The positions of the mathematical dislocations in a symmetric tilt boundary  $OO_1N$  are also shown.

of the two lattices in the vicinity of the  $O$ -point. This may be shown analytically by considering the relations between the two sets of lattice points when the origin is shifted to the  $O$ -lattice point  $x^{(0)}$ . Using eqn. (38.7), the relation

$$y - x^{(0)} = Sx - x^{(0)}$$

may be put in the form

$$y - x^{(0)} = S(x + u - x^{(0)}) \tag{38.9}$$

which is of the original form  $y = Sx$  but corresponding points of the two crystals are changed. Thus point  $y$  of the  $+$  crystal was originally derived from point  $x$  of the  $-$  crystal, but is now derived from the point  $x + u$  (measured from the first origin).

We can now make the transition from the theory of the  $O$ -lattice to a dislocation model of a boundary by separating the different  $O$ -elements (points, lines, or planes) by "cell

walls". The accumulating lattice misfit in any direction can be considered to be concentrated into discontinuities at the cell walls, and when a real interface in any given orientation is introduced into the space, the intersections of the cell walls with this interface become line discontinuities which are described as "mathematical dislocations". Each *O*-lattice point may now be considered as the origin for the relation  $y = Sx$  within its own cell; when a cell wall is crossed the relation between corresponding lattice points changes by the vector  $u$  which is the difference vector between two reference lattice points. The vector  $u$  satisfies both the Burgers circuit (or topological) definition of a dislocation and also the kinematical, or displacement, definition (Bollmann, 1970). Thus there is a formal equivalence with the Bilby theory, and this may be represented by the equation

$$b^{(1)} = -u = (S^{-1} - I)x^{(0)}, \quad (38.10)$$

where  $b^{(1)}$  are the possible Burgers vectors of primary mathematical dislocations in the interface. Secondary dislocations with Burgers vectors  $b^{(2)}$  are introduced below.

The mathematical dislocations are the discrete equivalents of the continuous distribution of dislocations in the Bilby theory, and we can now see that Bollmann's condition for the choice of  $S$  is equivalent to a maximization of the dislocation spacings or to a minimization of the net Burgers vector. For example, in the case of a pure rotation the three dimensional determinant  $|I - S^{-1}| = 0$  and the two-dimensional determinant

$$|I - A^{-1}| = 2(1 - \cos \theta) \quad (38.11)$$

is minimized by choosing that representation of  $A$  for which  $|\theta|$  is smallest. This is the same result as that obtained by minimizing  $|b^{\max}|$  in eqn. (36.18). However when more general forms are used, so that  $S$  is not a pure rotation even for grain boundaries, there is the possibility of still further reductions in the determinant, and in many cases the two-dimensional determinant may be made zero. The Bollmann condition is then inapplicable without additional arbitrary assumptions (Christian and Crocker, 1980).

We have noted above that if  $S$  represents an invariant plane strain, the *O*-lattice degenerates into an *O*-plane lattice. The cell walls of this lattice are parallel planes midway between the *O*-elements and an interface parallel to the *O*-planes thus intersects no cell walls and hence contains no mathematical dislocations. This description is self-consistent since the interface is fully coherent, i.e. it is the invariant plane of the deformation  $S$ .

The difficulty which now arises is that unless some restrictions are placed on  $S$  there are many zero dislocation descriptions. As previously noted, for example, all rational symmetrical tilt boundaries in cubic crystals may be represented as coherent twin boundaries and hence, by appropriate choice of  $S$ , as dislocation-free boundaries. If the Miller indices of the boundary are allowed to increase without limit, the region near  $\theta = 0^\circ$  becomes more and more populated with "coherent" twin relations. However, for low-angle boundaries with low densities of coincidence sites and very small spacing of lattice planes parallel to the boundary, the usual description (Section 36) of the boundary as an array of edge dislocations, which is obtained by considering  $S$  as a rotation, is physically more realistic. Thus the Bollmann criterion does not work in this case: the difficulty is exactly that mentioned on p. 352.

For high-angle grain boundaries, more general expressions for  $S$  are often appropriate, as is clearly shown by the existence of deformation twins where the simple shear  $S$  describes a physical correspondence, rather than just one of many possible mathematical representations. Even here however the Bollmann condition can only be used if combined with an assumption about which correspondences (or lattice reproducing shears  $U$ , eqn. (36.12)) are permissible: the decision is equivalent to a choice of which coincidence site relations have special properties. Thus Bollmann considered  $[1\bar{1}0]$  tilt boundaries in b.c.c. lattices and used two forms for  $A$ , one representing a rotation of  $\theta$  and the other relating corresponding unit cells as in the (112) twin. The second form gives

$$|I - A^{-1}| = 2 - \cos \theta - (5\sqrt{2}/4) \sin \theta, \quad (38.12)$$

and has a smaller absolute value than (38.11) for  $\theta \geq 29.5^\circ$ . Moreover, the determinant in (38.12) has zero values at  $\sim 50.5^\circ$  and  $70.5^\circ$ , corresponding to (332) ( $\Sigma = 11$ ) and (112) ( $\Sigma = 3$ ) twin boundaries. However, there is also a (114) ( $\Sigma = 9$ ) twin boundary at  $\sim 39^\circ$ , and there seems no reason why this also should not be regarded as a full coherent, zero dislocation boundary, although Bollmann chose not to do so.

For interphase boundaries, the possibilities of zero dislocation descriptions are remote because of the restrictive conditions needed to enable  $S$  to be described as an invariant plane strain (Section 9). However, there is an equivalent difficulty, since by choosing correspondences involving large shears and/or shuffles, the value of the Bollmann determinant may be reduced.

The element of arbitrariness in the primary dislocation description is also relevant to the secondary dislocation models of boundaries which deviate from ideal conditions. However, we first have to consider some other aspects of the primary model of discrete mathematical dislocations. The theory completely specifies their spacing in any given direction, but there is some ambiguity about the definitions of their positions. Bollmann assumes that the cell wall between, for example, the origin and the  $O$ -lattice point  $\mathbf{x}^{(0)}$  is defined by the condition that the magnitude of the relative displacement of corresponding points  $\mathbf{y}$  and  $\mathbf{x}$  is smaller within the cell around the origin than the relative displacement of new corresponding points  $\mathbf{y}$  and  $\mathbf{x} + \mathbf{u}$  produced by a change of origin to  $\mathbf{x}^{(0)}$ , but is larger outside this cell. This means that points on the cell wall may be defined by the condition

$$|\mathbf{y} - \mathbf{x}|^2 - |\mathbf{y} - \mathbf{x} - \mathbf{u}|^2 = 0 \quad (38.13)$$

or after some algebraic manipulation

$$\mathbf{u}' \mathbf{G} (\mathbf{y} - \mathbf{x}) - \frac{1}{2} \mathbf{u}' \mathbf{G} \mathbf{u} = 0, \quad (38.14)$$

where  $\mathbf{G}$  is the metric tensor. Finally, since  $\mathbf{y} - \mathbf{x} = (I - S^{-1})\mathbf{y}$ , the equation giving any point  $\mathbf{y}$  on the cell wall may be expressed as

$$\mathbf{u}' \mathbf{G} (I - S^{-1}) \mathbf{y} - \frac{1}{2} \mathbf{u}' \mathbf{G} \mathbf{u} = 0. \quad (38.15)$$

It follows from eqn. (37.7a) that eqn. (38.15) is satisfied by the point  $\mathbf{y} = \frac{1}{2}\mathbf{x}^{(0)}$ ; in general, the cell wall bisects, but is not necessarily perpendicular to, the line joining the two  $O$ -elements

which it separates. The whole procedure is equivalent to the mapping of the Wigner-Seitz-Slater type cells of the reference lattice into the real lattice.

When the  $O$ -lattice cell is large compared with those of the real lattices, the cell walls are spaced at many atom distances, and the structure of the interface, obtained by relaxation from the initial state we have just described, may correspond to arrays of physical dislocations. Unlike the mathematical dislocations, the positions of which are defined in the  $O$ -lattice, the physical dislocations cannot in general be uniformly spaced (cf p. 337). Thus in Fig. 8.18, mathematical edge dislocations corresponding to a symmetrical tilt boundary along  $OO_1N = (20, \bar{3}, 0)$  are at positions  $[\pm(n + \frac{1}{2})x_1^{(0)}, 0]$ , where  $n$  is any integer, and their mutual separation is  $\frac{1}{2}a \operatorname{cosec} \frac{1}{2}\theta$  along the interface or  $\frac{1}{2}a \cot \frac{1}{2}\theta = 10a/3$  projected along the  $[010]$  direction of the reference lattice. The physical dislocations may be placed on the  $(010)$  planes of the reference ( $-$ ) lattice if they are spaced at repeating intervals of  $3a$ ,  $3a$ , and  $4a$ . The centre of the physical dislocation should lie on the lattice plane nearest to a cell wall. For small  $O$ -lattice cells it may no longer be possible to describe the structure by physical dislocations. However, for both large and small  $O$ -lattice cells, any boundaries which contain a relatively high density of  $O$ -lattice points are likely to have lower energies than other boundaries. When the  $O$ -lattice becomes a lattice of  $O$ -lines, the cell walls consist of planes parallel to the invariant line, and any boundary which contains the invariant line will intersect only one type of cell wall. Thus the structure of an interface of this type may consist of a single set of parallel mathematical dislocations. This applies to all tilt boundaries and also to all martensitic interfaces. The formal theory of martensite crystallography includes a necessary condition that  $S$  represents an invariant line strain and that this invariant line lies in the interface.

Some complications arise when the plane with the invariant normal is not rational and thus contains few or no points of the reference lattice. This plane may be considered as a sub-space of the reference lattice, within which each lattice point gives rise to an image as an  $O$ -line. As the density of lattice points in this sub-space decreases, the distances between the  $O$ -lines increases and the Burgers vectors of the mathematical dislocations becomes very large. This scarcely seems a plausible physical model, and to overcome its limitations the effect of a relative translation of the two lattices may be considered. Bollmann shows that if the  $-$  lattice is displaced relative to the  $+$  lattice, the reference and  $O$ -lattice are also displaced. Let the displacement of the reference lattice be  $t$ ; then the displacement of the  $-$  lattice is  $S t$  and of the  $O$ -lattice is

$$t^{(0)} = (I - S^{-1})^{-1} t. \quad (38.16)$$

The geometrical structure of a three-dimensional  $O$ -lattice is preserved by this translation, but when the rank of  $(I - S^{-1})$  is only 2, the lattice of  $O$ -lines may disappear completely if  $t$  has a component normal to the plane of the reference lattice which has an invariant normal. Now consider again the situation when the plane with the invariant normal contains few lattice points. By allowing relative translations of the  $+$  and  $-$  lattices, these points of the reference lattice which are close to this plane may be projected on to it and a relatively dense lattice of  $O$ -lines may thus result. This means that the irrational plane of the reference lattice which generates the lattice of  $O$ -lines is replaced by a stepped series

of rational planes, which together generate the  $O$ -lattice. Physically, the transformation around each  $O$ -line is still described by  $S$  but in addition the lattices are allowed a small relative translation which may not be the same around different  $O$ -lines. This inhomogeneous transformation thus permits the periodicity of the interface structure to become much smaller, and the discontinuities less drastic, and thus it may be assumed that it is likely to give a lower energy.

In a similar way, for an  $O$ -plane lattice it is necessary to realize that the  $O$ -planes are the images of reference lattice points which lie along a line. If this line is not a rational lattice direction of the reference lattice adjacent lattice points may be projected on to it in order to give a reasonably dense lattice of  $O$ -planes.

The  $O$ -lattice theory may be regarded as a quantized version of the formal Bilby theory since it adds to this theory the requirement that the net Burgers vector  $\mathbf{b}$  of eqn. (38.2) is obtained by summation of individual mathematical or physical dislocation lines with lattice Burgers vectors. For the special case of grain boundaries we have also introduced the concept of secondary dislocations which represent the deviation from an ideal or coincidence site situation. In the same way an ideal or low energy interface between two lattices (if such exists) may in principle be used to give a secondary dislocation model of an interface which deviates from this ideal condition.

The theory of secondary dislocation arrays has been expressed in general form by Bollman, who introduced the concept of the " $O$ -2" lattice, which is an  $O$ -lattice formed between two  $O$ -lattices rather than between two real lattices. The two  $O$ -lattices correspond to the actual transformation  $S$  and the ideal or reference relation  $S_i$ , and are both defined in the usual way by eqn. (38.7). The relation between the two  $O$ -lattices is then

$$\mathbf{B} = \mathbf{S}^{-1} \mathbf{S}_i, \quad (38.17)$$

where  $\mathbf{B}$  specifies the deviation to be described by the secondary dislocation array. The  $O$ -2 lattice is now defined by an equation analogous to (38.7), namely

$$\mathbf{x}^{(02)} = (\mathbf{I} - \mathbf{B}^{-1})^{-1} \mathbf{u}^{(0)}, \quad (38.18)$$

where  $\mathbf{u}^{(0)}$  represents possible translations (Burgers vectors) in the ideal or reference  $O$ -lattice. Clearly, the translation repeat vectors of the  $O$ -lattice are possible vectors  $\mathbf{u}^{(0)}$ , just as crystal lattice repeat vectors are possible vectors  $\mathbf{u}$  in eqn. (38.7). However, the repeat vectors of the  $O$ -lattice only form a subset of the possible vectors  $\mathbf{u}^{(0)}$ , as will now be shown.

We pointed out in Section 36 that when a coincidence lattice exists there are a finite number of different internal coordinates  $\xi_i$  at the  $O$ -lattice points. The arrangement of crystal lattice points within one  $O$ -cell is referred to as a pattern element, and the combination of different pattern elements gives the complete pattern of lattice points. If the complete pattern is periodic, it consists of a finite number of pattern elements, this number being the same as the number of different values of  $\xi$ . Bollmann introduced the idea of the reduced  $O$ -lattice, which is defined as a single cell of lattice  $A$  within which the  $\xi_i$  values of the  $O$ -lattice points are plotted. Now consider a shift of the origin of the  $O$ -lattice to one of these internal points, and correspondingly a relative shift of the two crystal lattices. Such a displacement will obviously lead to conservation of the pattern as a whole since the displacement gives another pattern element. The lattice of all such displacements is called

the “complete pattern shift lattice” or “DSC lattice”; it is formed by an infinite repetition of the reduced  $O$ -lattice.

We have assumed that the existence of a periodic pattern implies a coincidence site lattice, but the pattern will remain periodic if the actual positions of the two sets of atoms are translated away from coincidence. Any such translation displaces the  $O$ -lattice by a vector which is not a repeat vector of the DSC lattice and there is no longer a reduced  $O$ -lattice point at the origin. The number of points in the reduced  $O$ -lattice, the number of pattern elements and the periodicity of the pattern, however, are all conserved. Clearly, the DSC lattice is also unchanged, and any translation through a vector of this lattice will reproduce the new pattern.

It is now clear that the vectors  $u^{(0)}$  of eqn. (38.18) may be any vectors of the DSC lattice since linear discontinuities with this displacement (Burgers vector) will reproduce the interface structure. However, the vectors  $u^{(0)}$  are referred to the  $O$ -lattice and may be transferred to the reference ( $-$ ) lattice through eqn. (38.16). Thus the possible Burgers vectors  $b^{(2)}$ , referred to the  $-$  lattice, are

$$b^{(2)} = (S^{-1} - I) u^{(0)} \quad (38.19)$$

which may be compared with (38.10). Combining (38.19) with (38.18) gives a new basic equation for the  $O$ -2 lattice which replaces eqn. (38.7) for the  $O$ -lattice as

$$(S^{-1} - I)(I - B^{-1}) x^{(02)} = b^{(2)}. \quad (38.20)$$

The  $O$ -2 lattice may be subdivided by cell walls in an exactly analogous manner to the  $O$ -lattice, and the intersection of each such wall with the interface then represents a mathematical “secondary” dislocation with a Burgers vector  $b^{(2)}$ . The positions of the cell walls are defined by an equation equivalent to (38.15), namely

$$b^{(2)'} G (I - S^{-1})(I - B^{-1}) \gamma - \frac{1}{2} b^{(2)'} G b^{(2)} = 0. \quad (38.21)$$

As already noted, the Burgers vectors  $b^{(2)}$  of the secondary dislocations can be markedly smaller than those of the primary dislocations.

As the theory of grain boundaries has developed, it has become clear that the concept of the DSC lattice is much more significant than that of the  $O$ -lattice. It may be most simply regarded as the lattice of vectors linking sites of one crystal lattice to those of the other when these two lattices are in a coincidence site orientation and position, and it is the coarsest lattice which contains both crystal lattices as a superlattice. Analytical methods of constructing the DSC lattice have been given by Loberg and Norden (1973), Grimmer (1973), Warrington and Grimmer (1974) and Warrington and Bollmann (1976). Many experimental observations of secondary dislocations have been made by thin film electron microscopy (a list is given by Dingley and Pond, 1979) especially following the development of techniques for making bicrystals by welding together single crystal films (Schober and Balluffi, 1971; Baluffi *et al.*, 1972). In some cases, contrast experiments have enabled the Burgers vectors to be identified as DSC lattice vectors; more usually this has to be deduced indirectly from measured grain misorientations and dislocation spacings. A distinction is often made between geometrically necessary or “intrinsic” grain boundary dislocations and “extrinsic” dislocations (called by Hirth and Balluffi, 1973, just grain boundary dislocations) which are present because of growth accidents, plastic deformation,

etc. Arrays of intrinsic dislocations as predicted have been observed for several near-coincidence boundaries, but there are often some extrinsic dislocations also present, and these sometimes have complex interactions with the intrinsic dislocations. Extrinsic dislocations arise, for example, from the dissociation of lattice dislocations entering the boundary, or from the operation of Bardeen–Herring type sources in the boundary.

As noted on p. 356, defects in a grain boundary may also occur at the intersection of two different translation states, and these have an effective Burgers vector given by the difference  $\Delta t$  of the translations. This is not a lattice vector of the DSC lattice, and not necessarily a rational fraction of such a vector, but the defects are generally called partial grain boundary dislocations. They may occur (a) at the intersection of two facets of different orientation and  $t$ , in which case they may be compared to stair-rod dislocations, or (b) at the junction of two regions in a planar facet. In the latter case, the two regions may be symmetry degenerate or one of them may be metastable with a higher energy than the other.

Finally, there is an important treatment of interface symmetry due to Pond and his collaborators. Pond and Bollman (1979) used the principles of colour symmetry to discuss the symmetry of two crystals joined at a planar interface and the theory was further developed by Pond and Vlachavas (1987) and Pond (1989). A rather similar development is due to Gratias and Thalal (1988) and Gratias and Portier (1983). This theory leads to a classification of all interface line defects which join together energetically degenerate regions which are either identical or equivalent variants of one another. Interface line defects producing displacements of this type may be called “perfect” in analogy with crystal dislocations which have lattice Burgers vectors. When two crystals are joined together in some specific orientation, the symmetry axes and planes of each of the crystals will be lost unless they coincide with symmetry axes and planes of the other crystal. This process is called dissymmetrization. New symmetry elements may be created in the case of grain boundaries only and these new elements relate the atoms of one crystal to the atoms of the other they are called antisymmetry elements or colour reversing elements.

Pond and Vlachavas developed their theory of “bicystallography” by defining dichromatic pattern and the dichromatic complex. The dichromatic pattern is obtained by colouring all the lattice points of one crystal white and those of the other structure black. The two lattices in the desired orientation relation are then allowed to interpenetrate and this defines the dichromatic pattern. The dichromatic complex takes account of the actual crystal structure not just the lattice and after colouring atoms of the two structures white and black respectively, they are allowed to interpenetrate to form the complex. Pond and Vlachavas deal with dissymmetrization in four stages, namely the reduced symmetry of the dichromatic pattern, followed by that of the dichromatic complex. The interface is introduced at the third stage by sectioning the dichromatic complex along a plane and removing the white atoms from one side of this plane and the black atoms from the other side. In the final stage the interatomic forces are switched on and the structure is allowed to relax to its minimum energy configuration.

The space group of the bicrystal is obtained from the intersection of the space groups of the two constituent crystals. A perfect defect in the interface must separate two regions of interface with the same structure, or with crystallographically equivalent structures which are simply variants of the same structure. Pond used the Principle of Symmetry Compensation (Shubnikov

and Koptsik, 1977) which shows that the destruction of a symmetry element always results in the creation of new variants. He showed that such defects are characterised by broken symmetries arising from the combined space group. He introduced a circuit operator which is a generalised Burgers circuit in which symmetry operations other than translation are included. The main defects raised by this treatment are dislocations with Burgers vectors given by vectors of the D.S.C. lattice but by using the properties of glide planes or screw-symmetric axes some previously unrecognised interfacial defects were discovered. Complete characterisation of the topological nature of defects in this way enabled not only classification of possible defects but also modelling of processes such as phase transformations. Interfacial line defects may have dislocation, disclination or dispiration character, just as in the interior of crystals, but a free surface can only have fault lines and steps.

The other general approach is the concept of the surface dislocation (see p. 366) as a convenient way of discussing the average structure of an interface. Clearly a surface dislocation, like a line dislocation, cannot end within a crystal. Three or more surface dislocations may, however, meet in a common line, which Bilby (1955) has called a surface dislocation node. There is then a compatibility condition on the individual surface dislocation tensors which may be compared with the equation for the conservation of Burgers vectors. Examples of surface dislocation nodes occur in the deformation patterns produced by indenting zinc single crystals (Li *et al.*, 1953), and in the structures produced by martensitic transformation in indium–thallium alloys (Basinski and Christian, 1956).

#### REFERENCES

- AMELINCKX, S. (1957) *Dislocations and Mechanical Properties of Crystals*, p. 3, Wiley, New York; (1958) *Phil. Mag.* **6**, 34.
- ASHBY, M. F., SPAEPEN, F. and WILLIAMS, S. (1978) *Acta metall.* **26**, 1647.
- AUST, K. T. and RUTTER, J. W. (1959) *Trans. Am. Min. (Metall.) Engrs.* **215**, 820.
- BALLUFFI, R. W., KOMEN, Y., and SCHOBER, T. (1972) *Surf. Sci.* **31**, 68.
- BASINSKI, Z. S. and CHRISTIAN, J. W. (1956) *Acta metall.* **4**, 371.
- BILBY, B. A. (1955) *Report of the Conference on Defects in Crystalline Solids*, p. 123, The Physical Society, London.
- BILBY, B. A. and SMITH, E. (1958) *Proc. Roy. Soc. A*, **236**, 481.
- BISHOP, G. H. and CHALMERS, B. (1968) *Scripta metall.* **2**, 133; (1971) *Phil. Mag.* **24**, 515.
- BOLLMANN, W. (1967) *Phil. Mag.* **16**, 363, 383; (1970) *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin.
- BORN, M. (1946) *Proc. Math. Phys. Soc. Egypt* **3**, 35.
- BRAGG, W. L. (1940) *Proc. Phys. Soc.* **52**, 54.
- BRANDON, D. G. (1966) *Acta metall.* **14**, 479.
- BRANDON, D. G., RALPH, B., RANGANATHAN, S., and WALD, M. S. (1964) *Acta metall.* **12**, 813.
- BRISTOWE, P. D. and CROCKER, A. G. (1978) *Phil. Mag.* **38**, 487.
- BROOKS, H. (1952) *Metal Interfaces*, p. 20, American Society for Metals, Cleveland, Ohio.
- BURGERS, J. M. (1939) *Proc. Akad. Sci. Amst.* **42**, 293.
- CARRINGTON, W., HALE, K. F. and MCLEAN, D. (1960) *Proc. Roy. Soc. A*, **259**, 203.
- CHALMERS, B. and GLEITER, H. (1971) *Phil. Mag.* **23**, 1541.
- CHAUDHARI, P. and CHARBNAU, H. (1972) *Surf. Sci.* **31**, 104.
- CHRISTIAN, J. W. (1962) *The Decomposition of Austenite by Diffusional Processes*, p. 371, Wiley, New York.
- CHRISTIAN, J. W. and CROCKER, A. G. (1980) In *Dislocations in Solids* (ed. F. R. N. Nabarro), vol. 3, p. 165, North-Holland, Amsterdam.
- COUCHMAN, P. R., JESSER, W. A., and KUHLMANN-WILSDORF, D. L. (1972) *Surf. Sci.* **33**, 429.
- CROCKER, A. G. and FARIDI, B. A. (1980) *Acta metall.* **28**, 549.

- DAHL, R. E., Jr., BEELER, J. R., and BOURQUIN, R. D. (1972) *Interatomic Potentials and Simulation of Lattice Defects*, p. 673, Plenum Press, New York and London.
- DINGLEY, D. J. and POND, R. C. (1979) *Acta metall.* **27**, 667.
- FORTES, M. A. (1972) *Revista de física Química e Engenharia* **4A**, 7.
- FRANK, F. C. (1950) *Symposium on the Plastic Deformation of Crystalline Solids*, p. 150, Office of Naval Research, Pittsburgh Pennsylvania; (1955) *Report of the Conference on Defects in Crystalline Solids*, p. 159, The Physical Society, London.
- GRATIAS, D. and PORTIER, R. (1983) *J. de Phys.* **43**, C6-15.
- GRATIAS, D. and THALAL, A. (1988) *Phil. Mag. Letts.* **57**, 63.
- GRIMMER, H. (1973) *Scripta metall.* **7**, 1295.
- GUYOT, P. and SIMON, J. P. (1976) *Phys. stat. sol. (a)* **38**, 207.
- HASSON, G., BOOS, J. Y., HERBEUVAL, I., BISCONDI, M., and GOUX, C. (1972) *Surf. Sci.* **31**, 115.
- HERRING, C. (1951) *The Physics of Powder Metallurgy*, p. 143, McGraw-Hill, New York; (1953) *Structure and Properties of Solid Surfaces*, p. 5, University of Chicago, Chicago, Illinois.
- HIRTH, J. P. and BALLUFFI, R. W. (1973) *Acta metall.* **7**, 929.
- INGLE, K. W. and CROCKER, A. G. (1980) *Phil. Mag.* **A 41**, 713.
- KRONBERG, M. L. and WILSON, F. H. (1949) *Trans. Am. Inst. Min. (Metall.) Engrs.* **185**, 50.
- LI, J. C. M. (1972) *Surf. Sci.* **31**, 12.
- LI, C. H., EDWARDS, E. H., WASHBURN, J., and PARKER, E. R. (1953) *Acta metall.* **1**, 223.
- LOBERG, B. and NORDEN, H. (1973) *Acta metall.* **21**, 213.
- LODGE, K. W. and FLETCHER, N. H. (1975) *Phil. Mag.* **31**, 529.
- LOMER, W. M. and NYE, J. F. (1952) *Proc. R. Soc. A*, **212**, 516.
- MEIJERING, J. L. (1953) *Acta metall.* **1**, 607.
- MOTT, N. F. (1948) *Proc. Phys. Soc.* **60**, 391.
- POND, R. C. (1977) *Proc. Roy. Soc. A* **357**, 471; (1979) *Phil. Trans. Roy. Soc. A* **292**, 449; (1989) *Dislocations in Solids*, Ed. F.R.N. NABARRO, Vol 8 p.165 North Holland, Amsterdam
- POND, R. C. and BOLLMAN, W. (1979) *Phil. Trans. Roy. Soc. London* **A292**, 449.
- POND, R. C. and SMITH, D. A. (1974) *Can. Metall. Quart.* **13**, 39.
- POND, R. C., SMITH, D. A. and VITEK, V. (1979) *Acta metall.* **27**, 235.
- POND, R. C. and VITEK, V. (1977) *Proc. Roy. Soc. A* **357**, 453.
- POND, R. C. and VLACHAVAS, D.S. (1987) *Proc. Roy. Soc. A* **386**, 95.
- PUMPHREY, P. H., MALIS, T. F. and GLEITER, H. (1976) *Phil. Mag.* **34**, 227.
- RANGANATHAN, S. (1966) *Acta Crystallogr.* **21**, 197.
- READ, W. T. (1953) *Dislocations in Crystals*, McGraw-Hill, New York.
- READ, W. T. and SHOCKLEY, W. (1950) *Phys. Rev.* **78**, 275; (1952) *Imperfections in Nearly Perfect Crystals*, p. 77, Wiley, New York.
- SCHOBER, T. and BALLUFFI, R. W. (1971) *Phil. Mag.* **24**, 165, 469.
- SHOCKLEY, W. and READ, W. T. (1949) *Phys. Rev.* **75**, 692.
- SHUBNIKOV, A.V. and KOPTSIK V.A. (1977) *Symmetry in Science and Art*, Plenum Press, New York.
- SHUTTLEWORTH, R. (1950) *Proc. Phys. Soc. A*, **63**, 444.
- SMITH, C. S. (1952) *Metal Interfaces*, p. 65, American Society for Metals, Cleveland, Ohio; (1953) *Acta metall.* **1**, 244.
- SMITH, D. A., VITEK, V. and POND, R. C. (1977) *Acta metall.* **25**, 475.
- SMOLUCHOWSKI, R. (1952) *Phys. Rev.* **87**, 482.
- SUTTON, A. P. and VITEK, V. (1980) *Scripta metall.* **14**, 129
- VAN DER MERWE, J. H. (1950) *Proc. Phys. Soc. A*, **63**, 613.
- VITEK, V. (1970) *Scripta metall.* **4**, 725.
- WARRINGTON, D. H. and BUFALINI, P. (1971) *Scripta metall.* **5**, 771.
- WARRINGTON, D. H. and BOLLMANN, W. (1976) *Phil. Mag.* **25**, 771.
- WARRINGTON, D. H. and GRIMMER, H. (1974) *Phil. Mag.* **30**, 461.
- WEINS, M. J. (1972) *Surf. Sci.* **31**, 138.

## CHAPTER 9

### *Diffusion in the Solid State*

#### 39. MECHANISM OF ATOMIC MIGRATION

It is usual to begin a discussion of diffusion with a statement of Fick's law, which is the analogue for material flow of Fourier's law for heat flow by conduction. Although it is not strictly valid, and has had to be modified in the formal theory, Fick's equation retains its significance since experimental results are almost always presented in the form of diffusion coefficients defined by reference to it. We shall find it simpler, however, to approach the subject in a more fundamental manner by considering first the ways in which atoms may move through the crystal. A general account of diffusion can be developed without reference to the actual mechanism of atomic migration, but the more detailed theories depend on this mechanism, and may readily be linked to our previous discussion of point defects.

Consider an assembly containing two or more kinds of different atoms. "Chemical" diffusion takes place when this assembly is not in equilibrium with respect to distribution of the different atomic species, so that there is, for example, a displacement of  $B$  atoms relative to  $A$  atoms. "Tracer" diffusion presupposes a situation in which the assembly is in equilibrium, except for the distribution of two components which differ from each other in an insignificant manner (so far as the diffusion problem is concerned). These two components must thus be virtually identical in chemical properties, but must be "labelled" in some way, e.g. by their nuclear properties. Tracer diffusion is, in fact, generally investigated by setting up a concentration gradient of a radioactive isotope of one of the components of an otherwise homogeneous assembly. Self-diffusion usually means tracer diffusion of an isotope of a pure element, or of the solvent atoms of a dilute solution.

Consider the chemical diffusion of an atomic species,  $B$ . The  $B$  atoms may share a set of sites with the other components, forming a substitutional solid solution, or they may partially occupy a separate set of sites, so that  $B$  is an interstitial solute. In the latter case, the mechanism by which  $B$  atoms diffuse is fairly certain, and needs little comment. In all interstitial solutions, the fraction of possible solute sites which are actually occupied by atoms is quite small, and most of these sites are vacant. An interstitial atom can then migrate by jumping from one such site to a neighbouring vacant site. This process is usually treated by the chemical rate theory of Chapter 3, there being a saddle point of free energy at some intermediate configuration which is critical for the attempt at a jump to succeed. Interstitial diffusion is shown diagrammatically in Fig. 9.1(a). We note that the  $A$  atoms form

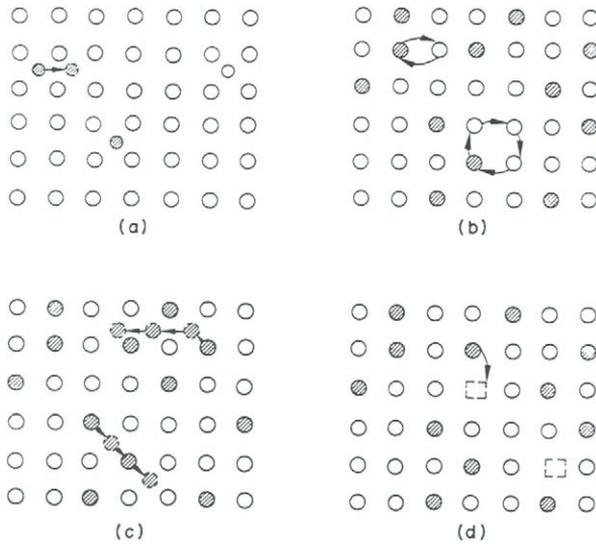


FIG. 9.1. Possible atomic mechanisms for diffusion. *A* atoms; open circles. *B* atoms; filled circles. (a) Solute diffusion in interstitial solid solution. (b) Place exchange and ring diffusion. (c) Interstitial and interstitialcy diffusion. (d) Vacancy diffusion.

a rigid frame of reference, relative to which the migration of the *B* atoms may be studied.

For substitutional solid solutions, the atomic mechanism is less certain. There are three basic possibilities for the relative displacement of *A* and *B* atoms occupying a common set of sites. The first of these is shown in Fig. 9.1(b), and involves the direct interchange of two atoms on neighbouring sites. If this interchange is imagined to occur by the rotation of the two atoms about their common centre of separation, it is obvious that considerable distortion of the surrounding structure will be required. For this reason, the probability of such a direct interchange is believed to be very low in metallic crystals. As a result of the process, both *A* and *B* atoms are displaced (in opposite directions) relative to the remaining atoms.

The second basic mechanism is the movement of atoms through the interstices of the lattice, as for interstitial solid solutions. An individual atom migrates by leaving a normal site and successively occupying a series of interstitial sites, until it meets a vacancy and returns to the normal sites. Such migration will be energetically unfavourable unless the structure is very "open". The interstitial mechanism is illustrated in Fig. 9.1(c) which also shows a variation suggested by Seitz (1950a, b). In Seitz's "interstitialcy" mechanism, the interstitial atom does not "squeeze past" the atoms on normal sites, but occupies the next normal site, "pushing" the atom on this site into an interstitial position, and so on. We have already discussed this mechanism in Section 17 for the case of self-interstitials.

Figure 9.1(d) shows the third mechanism for atomic migration, which depends on the presence of vacancies in the structure. An atom next to a vacant site may jump into that site, thus in effect changing places with the vacancy. The relative motion of *A* and *B* atoms results from the migration of vacancies through the lattice, and diffusion by vacancies is

in some of its effects intermediate between the interstitial and place exchange mechanisms. If only  $B$  atoms could change places with a vacancy, the motion of  $B$  would take place with reference to an unchanging lattice frame of  $A$  atoms. On the other hand, if the probabilities of  $A$  and  $B$  atoms jumping into a neighbouring vacant site were equal, the average displacement of the two kinds of atom would be equal and opposite, as in place exchange.

The above suggestions were once thought to cover all possibilities for the atomic mechanism of diffusion in the solid state. This is true so long as processes involving one or two atoms are considered, but it is now recognized that the co-operative motion of a larger number of atoms is also possible. The first type of co-operative process, suggested by Zener (1950) is an extension of the place exchange mechanism, and is called ring diffusion. A small number of atoms  $n$  move together in a ring, thus enabling the distortion to be spread over a larger volume, and reducing the energy. As the number of atoms in the ring increases, the activation energy first falls steeply, and then rises slowly; the minimum for a b.c.c. structure probably corresponds to  $n = 4$ . Ring diffusion is illustrated in Fig. 9.1(b).

On p. 128, we considered the possibility that an interstitial defect may spread its distortion along a line to form a crowdion. This suggestion was first made by Paneth (1950) as a possible diffusion mechanism in the alkali metals, and the motion of a crowdion may be regarded as a co-operative version of the interstitial mechanism. In the b.c.c. structure, crowdions will form (if at all) along  $\langle 111 \rangle$  directions. In general, crowdions should be able to migrate readily along nearest neighbour axes in the crystal, but cannot easily turn corners.

An additional mechanism for diffusion depends on the migration of vacancy pairs rather than of single vacancies. As discussed in Section 17, the divacancies migrate more readily than single vacancies, but under conditions of thermal equilibrium the concentration of divacancies is much smaller than that of single vacancies. This implies that divacancies contribute relatively little to the equilibrium diffusion rate of most metals, although this conclusion depends sensitively on the value of the divacancy binding energy  $2\Delta h_{\square} - \Delta h_{\square\square}$ . The divacancy contribution increases with increasing temperature, and with a high binding energy it may be  $\sim 10\%$  of the total diffusion at temperatures near the melting point. Moreover, pair or higher cluster migration may become important when there are large numbers of excess vacancies present, e.g. in a quenched specimen. The divacancy binding energy is again important, since it determines the average number of jumps made by a divacancy before it dissociates again into single vacancies. It is perhaps worth noting here that non-equilibrium concentrations of point defects may be very effective in producing enhanced atomic migration over short distances, but are unlikely to contribute much to long-range diffusion effects, since the defects themselves anneal out so rapidly. Enhanced diffusion will be obtained, however, if a high defect concentration can be maintained throughout the diffusion anneal by external means.

A co-operative diffusion mechanism suggested by Nachtrieb and Handler (1954) is based on the idea (p. 128) that a "relaxation" or local melting of the structure occurs in the immediate vicinity of a vacancy. The disordering of a small number of atoms, amongst which the volume of the vacancy is distributed, leads to rapid rearrangements by processes equivalent to liquid diffusion. The disordered volume itself is supposed to move through the

structure by local "melting" and "freezing" of atoms at its boundaries, one or two at a time. If its configuration remained constant during this migration, the whole process would be equivalent to vacancy diffusion, but it is supposed that rearrangements within the disordered volume occur at about the same rate as the movement of the volume itself. According to this theory, the various atomic mechanisms listed above may thus all be involved to some extent in the actual diffusion process.

In the above discussion, we have confined ourselves mainly to geometrical possibilities, and mentioned energies only occasionally. For any particular metal, we expect one of these possibilities to be favoured at the expense of the others, and thus to be the dominant mechanism for diffusion. We briefly repeat the reasons for this. The atoms, though vibrating about their equilibrium positions, spend most of their time in the vicinity of an atomic site, or minimum in the energy field. In order to move away from a given position, an atom must temporarily acquire an energy much greater than the mean thermal energy of vibration. The activation energy depends on the process concerned, and since the rate of the process is inversely proportional to the exponential of the corresponding free energy of activation, the mechanism with the lowest activation energy will probably be much more important than any other mechanism. Nevertheless, the accuracy of diffusion measurements is now often sufficient to permit contributions from the second most important process to be recognized and analysed.

The 'dominant' mechanism need not be that of lowest activation energy if different geometries exist to compensate for the different probabilities of individual atom movements. As we shall discuss later, for example, the activation energy for grain boundary migration is normally much lower than that for migration through the lattice. Nevertheless, the lattice makes the greater contribution to the overall diffusion rate at high temperatures because so few of the atoms are in grain boundary regions. As the temperature is lowered, the effect of the lower activation energy for boundary motion becomes progressively more important, and at sufficiently low temperatures virtually the whole of the diffusion may be along the boundaries.

Various methods have been used to deduce the principal mechanism of diffusion in particular metals or alloys. As discussed in Chapter 5, theoretical estimates of the activation energy are subject to severe limitations, but they may nevertheless sometimes enable meaningful distinctions to be made. Such estimates may also be compared with experimental activation energies, obtained from the measured variations of diffusion rates with temperature, and with the energies of formation and motion of the various atomic defects discussed in Section 17. A related method is to measure the pressure dependence of the diffusion coefficient, which defines an activation volume, and to compare this with predictions for the various models. For self-diffusion in pure metals, it has been suggested that the magnitude of the measured entropy of activation also provides a useful indication of mechanism.

Another group of methods depends on differences which arise from the geometry of the defects. For example, h.c.p. metals are anisotropic in diffusion properties and the ratio of the diffusion coefficients parallel and perpendicular to the basal plane is partly determined by entirely geometrical factors which differ with the mechanism. This has been used by Shirn *et al.* (1953) to deduce that vacancy diffusion is probably dominant in zinc, cadmium, and thallium. For cubic metals, we shall define later a quantity known as the correlation

factor which has purely geometrical values in the case of tracer self-diffusion. This factor may be derived in various ways, e.g. from measurements of the self-diffusion coefficients of different isotopes of the same element, and the results used to differentiate between alternative possible mechanisms. An older method depends on the distinction made above between the interchange and interstitial mechanisms. If the lattice structure is regarded as a rigid framework, diffusion by an interchange method will not result in the displacement of the atoms relative to this framework. Diffusion by the other two methods can, however, lead to a mass flow of atoms, since there is no need for the motion of the *B* atoms to be balanced by an equal and opposite motion of *A* atoms. The experimental observation of this phenomenon is now known as the Kirkendall effect, and first became widely known after the work of Smigelskas and Kirkendall (1947) on diffusion in copper-zinc alloys. The mass motion was detected by inserting inert markers (molybdenum wires) at the original interfaces of a copper-brass-copper diffusion couple. At the end of the experiment, the separation of the wires was found to have decreased, and the magnitude of the change was too great for it to be attributed simply to the different atomic volumes of the diffusing species. If the wires, which take no part in the diffusion process, are assumed to be fixed to the lattice, the effect can only be explained by assuming that the number of zinc atoms which diffuse out of the brass is greater than the number of copper atoms which diffuse into it.

In a large number of subsequent investigations, it has been shown that the Kirkendall effect must usually be ascribed to unequal rates of diffusion of the two species through the lattice and not to secondary causes, such as transfer in the vapour phase. Thus, in the above experiment, the zinc atoms could diffuse interstitially from the brass, leaving vacancies, and take up normal sites in the copper. Alternatively, the zinc atoms may diffuse via vacancies, the net current of atoms being balanced by an opposite current of vacancies into the brass. The vacancies could coagulate together, forming macroscopic holes, or could diffuse to dislocations, with a resultant shrinking of the structure. The extra atoms in the copper would similarly cause an expansion of this region.

The results of both experimental and theoretical methods of determining the diffusion mechanism have been reviewed by many authors, e.g. Seitz (1950a, b), Huntington (1951), Lazarus (1960), Shewmon (1963), Howard and Lidiard (1964), Peterson (1968), and Seeger (1972). Seeger concludes that in all metals investigated in detail, the dominant mechanism of self-diffusion is the migration of monovacancies, and the migration of divacancies is the second most important process. The metals considered by Seeger were all f.c.c. and the mechanism in b.c.c. metals has always been more uncertain, although the existence of Kirkendall effects in many cases gives a strong indication that vacancy diffusion is dominant here also. However, there are certain anomalies in the diffusion behaviour of many b.c.c. metals which are not yet fully explained, and these will be described later. There are also some special situations; for example, solute diffusion of the noble metals in germanium, silicon, lead, tin, indium, and thallium appears to involve a combination of vacancy and interstitial mechanisms with the interstitial process dominant (Frank and Turnbull, 1956; Dyson *et al.*, 1966, 1967, 1968).

## 40. STATISTICAL BASIS OF DIFFUSION: FICK'S LAW

A typical diffusion experiment consists in bringing two homogeneous specimens into contact at an interface. The specimens may be different pure components, or may contain the same components in different concentrations. After a period of time, it is found that the sharp change of composition has been replaced by a more gradual change, extending for some distance on either side of the original interface. A diffusion process of this type, which tends to produce a homogeneous assembly, is easily visualized. If the probability of an atomic migration were independent of direction, i.e. of the neighbours of the atom which moves, then diffusion would be merely the statistical result of a large number of chance migrations. This is often referred to as a "random walk" problem. The net statistical flow of any component would obviously be from regions of high concentration of that component to regions of low concentration.

In practice, the probability of atomic migration usually depends both on the type of atom and the direction of motion, and simple statistical flow does not occur. However, it is often convenient to divide the net atomic movements into a random flow and a superimposed drift velocity. The driving force for the random flow is the change in the configurational entropy of the assembly, whilst the drift velocity may be attributed to internal or external stresses, thermal gradients, or electrical fields, etc. A drift velocity in non-ideal solutions results from gradients of chemical potential.

The simple kinetic theory which we shall develop first in this section shows that random flow leads to a linear relation between flux  $I$  and concentration gradient  $\nabla c$ , and this is used to define a diffusion coefficient. The drift velocity is treated separately unless it is also proportional to  $\nabla c$ , in which case it may be combined with the statistical flow to define a different diffusion coefficient. An example is the important case of chemical diffusion in binary alloys. Since the terminology used in different books is not always identical, we find it convenient here to define the various diffusion coefficients by reference to experimental measurements. The three main types are tracer diffusion coefficients, intrinsic diffusion coefficients, and chemical or interdiffusion coefficients (Peterson, 1968).

Tracer-diffusion coefficients are measured from the atomic flux of an isotope which is present in low concentration in an otherwise homogeneous matrix. If the tracer is an isotope  $A_+$  of the only or major component  $A$ , the coefficient  $D_{A_+}$  is the tracer self-diffusion coefficient, whereas if the isotope  $B^+$  is a minor component,  $D_{B^+}$  is a tracer impurity or solute diffusion coefficient; in concentrated solutions these terms are not appropriate. Tracer self-diffusion by means of a vacancy mechanism is closely related to the motion of the vacancies, which may be used to define a vacancy diffusion coefficient. The vacancy motion in a pure element may be treated to a sufficient approximation as a random walk, but we shall show that the tracer jumps are not strictly random. The tracer self-diffusion coefficient is often called simply the self-diffusion coefficient, but experiments on the mass flow in a pure element subjected to a temperature gradient or an electrical field, etc., lead to the definition of another coefficient, the macroscopic or transport self-diffusion coefficient (Seeger, 1972). The atomic displacements in macroscopic self-diffusion are equivalent to a

true random walk and the ratio of the tracer to the macroscopic self-diffusion coefficient is called the correlation factor (see Section 41).

The intrinsic diffusion coefficient  $D_A$  is used to describe the flux of  $A$  atoms in an alloy containing concentration gradients of  $A$  and hence of another component  $B$ . This is the type of diffusion which is much more difficult to discuss theoretically than is tracer diffusion. In a binary alloy, intrinsic diffusion coefficients  $D_A$  and  $D_B$  may be defined, but diffusion in a ternary alloy cannot be treated in this way unless the third component  $C$  is uniform in concentration. The chemical or interdiffusion coefficient  $D_{\text{chem}}$  also describes the diffusion process in a binary alloy with a concentration gradient, and it is used to describe the rate of mixing (or unmixing!) of the two species. It follows that  $D_{\text{chem}}$  may be expressed in terms of  $D_A$  and  $D_B$ .

The usual diffusion experiments involve net flow down the concentration gradient, even though this flux is not the result of purely random migrations. These investigations, in which both parts of the specimen are initially at equilibrium, do not correspond to the conditions which are often important in transformations, such as diffusion in a supersaturated solid solution. In certain circumstances, diffusion may take place against the direction of purely statistical flow, thus producing local regions which differ in concentration. This is sometimes called "uphill diffusion", and will occur when the probability of migration against the concentration gradient is large enough to affect the flow more than the greater number of atoms which are available for motion down this gradient.

The above description emphasizes the distinction between the individual migrations of the atoms and the net resultant flow. If the assembly as a whole is not in thermodynamic equilibrium, diffusion will redistribute the atoms in the direction of this equilibrium. A given atom moves from its position as a result of a fluctuation in the local thermal energy. There is no restriction on the direction in which it may move, and all that we can do is to calculate the relative probabilities of its moving in different directions. If these probabilities are not equal, the assembly is not in an equilibrium state, and the net result of a large number of atoms undergoing a large number of displacements will be to provide a net flow in some direction. As already stated, it is this net flow which we describe as diffusion, and it is misleading, and in general incorrect, to use the term for the motion of atoms over distances of the order of interatomic distances. Diffusion rates are often interpreted in terms of the velocities of individual atoms, expressed as the product of an average diffusion "force" and a "mobility" (compare the theory of Chapter 4). This formal approach can be very useful, but the true statistical nature of the atom movements must always be kept in mind.

Let us consider a simplified situation in which there is net flow in only one direction of a crystal of a binary alloy. We use the labels 1 and 2 for two neighbouring planes, distance  $d$  apart, perpendicular to the direction of flow, and we assume that the elementary diffusion process consists of atomic jumps between positions on planes of this type. The numbers of  $A$  atoms on unit areas of the two planes may be written  $c_{A1}d$ ,  $c_{A2}d$ , where  $c_{A1}$ ,  $c_{A2}$  are the concentrations of  $A$  atoms per unit volume at the two planes. If the probability per unit time that an  $A$  atom on plane 1 will jump to plane 2 is written  $\pi_{A,12}$ , and that of the reverse jump is  $\pi_{A,21}$ , the net flow of  $A$  atoms between the two planes is

$$I_A = c_{A1}d\pi_{A,12} - c_{A2}d\pi_{A,21} \quad (40.1)$$

The frequency of jumping,  $\pi$ , depends on the chemical environment of the atom concerned. If all atoms are chemically identical, the frequency of jumping for a vacancy or interstitial mechanism may be assumed to be constant, and we write

$$\begin{aligned} \pi_{A, 12} &= \pi_{A, 21} = \pi_A, \\ I_A &= \pi_A d(c_{A1} - c_{A2}) = -\pi_A d^2 \frac{\partial c_A}{\partial x} = -D_A \frac{\partial c_A}{\partial x}, \end{aligned} \quad (40.2)$$

where the  $x$  axis has been taken perpendicular to the planes 1 and 2. This is the simplest form of Fick's law, with a steady flow of atoms proportional to the gradient of concentration, expressed in atoms per unit volume. The diffusion coefficient,  $D_A = \pi_A d^2$ , is constant, and has units  $\text{m}^2 \text{s}^{-1}$ . We shall also have

$$I_B = -\pi_B d^2 \frac{\partial c_B}{\partial x} = -D_B \frac{\partial c_B}{\partial x}. \quad (40.3)$$

Since  $\pi_A = \pi_B$  if all atoms are equivalent, and  $\partial c_A / \partial x = -\partial c_B / \partial x$ ,  $I_A = -I_B$ , and the diffusion coefficients are equal. This result applies, at least approximately, to chemical diffusion in an ideal solution, or to the diffusion of a radioactive isotope.

The frequency  $\pi_A$  is related to the frequency  $k_A$  with which an  $A$  atom in any site moves to a neighbouring vacant site. If there are  $z$  nearest sites to which an atom on plane 1 may jump, and a fraction  $1/j$  of these lie on the plane 2, then

$$\pi_A = Pz k_A / j, \quad (40.4)$$

where  $P$  expresses the probability that a given site is vacant, and is thus effectively unity for interstitial diffusion, and equal to  $x_{\square}$  for vacancy diffusion. The rate constant  $k_A$  is expressed in terms of an atomic frequency factor and a free energy of activation by means of the theory of Chapter 3.

The simple assumption that the atomic jumps take place between two neighbouring planes may readily be generalized to the case where the composition gradient has arbitrary direction and atomic jumps of different effective length along this direction are possible. Let  $\Gamma_A = Pz k_A$  be the total number of jumps made by an  $A$  atom in unit time (see p. 137 for the analogous problem of vacancy jumps). If successive jumps represent vector displacements  $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \text{etc.}$ , the total displacement after a large number of jumps  $n = \Gamma_A t$  will be a vector  $\mathbf{S} = \sum_{i=1}^n \mathbf{s}_i$ . We cannot predict  $\mathbf{S}$  for any given atom, but we can find a mean value,  $\overline{S^2}$ , of the square of the distance travelled by a large number of identical atoms. A simple kinetic argument, essentially equivalent to that above, then shows that the diffusion coefficient for a cubic material is given by

$$D_A = \overline{S^2} / 6t. \quad (40.5)$$

This is sometimes called the Einstein-Smoluchowski relation and was derived by Einstein (1905).

Expanding  $\mathbf{S} = \sum \mathbf{s}_i$ , we find that the mean value of  $S^2$  will be given by

$$\overline{S^2} = \sum_{i=1}^n \overline{s_i^2} + 2 \sum_{i=1}^{n-j} \sum_{j=1}^{n-1} \overline{\mathbf{s}_i \cdot \mathbf{s}_{i+j}} \quad (40.6)$$

We consider only the case when all elementary jumps are of the same length; this is believed to cover diffusion in both f.c.c. and b.c.c. structures. We let  $|\mathbf{s}_i| = r_1$ , the nearest neighbour distance for the sites concerned in the diffusion process. Then (40.6) becomes

$$\overline{S^2} = nr_1^2 + 2r_1^2 \sum_{j=1}^{n-1} (n-j) \overline{\cos \theta_j}, \quad (40.7)$$

where  $\overline{\cos \theta_j}$  is the average value of the cosine of the angle between the directions of the  $i$ th and  $j$ th jumps. If the motion of the atoms considered is a true random walk, in which there is no correlation between the directions of the individual jumps made by an atom,  $\overline{\cos \theta_j}$  must be zero for all values of  $j$ . We thus have the expression

$$D_A = \Gamma_A r_1^2 / 6, \quad (40.8)$$

which may be compared with that given by (40.2) and (40.4)

$$D_A = \Gamma_A d^2 / j. \quad (40.9)$$

The two equations are equivalent if the factor  $(1/j)$  is interpreted as the fractional number of jumps which are effectively along the composition gradient, and  $d$  is the average distance covered by each such jump. If the sites for the diffusing atoms have cubic symmetry, the flow of atoms depends only on the concentration gradient, and the diffusion coefficient is a constant. In crystals of lower symmetry, there will be non-equivalent sites to which an atom may move at different rates. The diffusion coefficient is then a second-order tensor, relating the two vector quantities  $\mathbf{I}_A$  and  $\text{grad } c$ . Instead of eqn. (40.5) we have three equations of form

$$D_i = \overline{S_i^2} / 2t$$

relating the principal values of the tensor  $\mathbf{D}$  to the mean square displacements along the principal axes.

As we shall see in a later section (p. 396), the assumption that  $\overline{\cos \theta_j} = 0$  is not valid, even for tracer self-diffusion by a vacancy mechanism. In practice, the most important example in which diffusion may be treated as a random walk problem is probably the diffusion of interstitial solute atoms. In the f.c.c. structure, the interstitial positions are at the body centre and the centres of the edges of the cubic unit cell, each site having twelve nearest neighbour sites. From eqns. (40.8) or (40.9) we find

$$D = \left(\frac{1}{12}\right) \Gamma_A a^2 = k_A a^2, \quad (40.10)$$

where  $a$  is the cube edge. Similarly, the interstitial sites in the b.c.c. structure are at the centres of the faces and edges of the unit cube, each site having four nearest neighbours. In this case

$$D = \left(\frac{1}{24}\right) \Gamma_A a^2 = \left(\frac{1}{8}\right) k_A a^2. \quad (40.11)$$

In deriving (40.10) and (40.11), we have assumed that all the interstitial sites next to an interstitial atom are vacant, so that  $P$  of eqn. (40.4) is unity. This will be true in dilute solution, and it is also the reason why we are able to treat interstitial diffusion as a random process, since the jump frequency will be almost constant throughout the specimen.

Returning to eqn. (40.1) we can assume that for a place exchange mechanism, the probability of an  $A$  atom jumping to plane 2 from plane 1 will be proportional to the number of  $B$  atoms on this plane, i.e. to  $c_{B2} = c - c_{A2}$ . We can thus write

$$\pi_{A, 12} = \pi'_A(c - c_{A2})/c, \quad \pi_{A, 21} = \pi'_A(c - c_{A1})/c,$$

and 
$$I_A = \pi'_A d(c_{A1} - c_{A2}), \tag{40.12}$$

which is of the same form as (40.2) with the diffusion coefficient  $D_A = \pi'_A d^2$ . Once again,  $I_A = -I_B$ , and  $D$  is independent of composition.

In deriving these results we have considered only atoms on the two neighbouring planes. We therefore obtain the same equations if we replace the assumption that all atoms are chemically equivalent by the slightly less restrictive condition that the jump frequencies  $\pi$  depend only on the average concentration near the jumping atom. If we write  $c_A = \frac{1}{2}(c_{A1} + c_{A2})$ , the constants  $\pi_A, \pi'_A$  now become parameters  $\pi_A(c_A), \pi'_A(c_A)$  which vary with composition. Fick's law now has to be written

$$I_A = -\partial(Dc_A)/\partial x. \tag{40.13}$$

In practice,  $D$  is nearly always a function of composition, and it is more realistic to express the difference between  $\pi_{A, 12}$  and  $\pi_{A, 21}$  in terms of the variation of  $\pi_A$  with  $c_{A1}$  and  $c_{A2}$ . We discuss this more satisfactory kinetic theory in Section 43.

The form of Fick's law given in eqn. (40.13) is inconvenient, since stationary flow can seldom be obtained in practice. The differential form of the law is obtained by considering the rate of accumulation of atoms in a small volume. For one-dimensional diffusion

$$\frac{\partial c_A}{\partial t} = \frac{\partial}{\partial x} \left\{ D \frac{\partial c_A}{\partial x} \right\}$$

and the general equation for three-dimensional diffusion is

$$\frac{\partial c_A}{\partial t} = \frac{\partial}{\partial x_1} \left\{ D_1 \frac{\partial c_A}{\partial x_1} \right\} + \frac{\partial}{\partial x_2} \left\{ D_2 \frac{\partial c_A}{\partial x_2} \right\} + \frac{\partial}{\partial x_3} \left\{ D_3 \frac{\partial c_A}{\partial x_3} \right\}, \tag{40.14}$$

where the three diffusion coefficients allow for anisotropy of diffusion in non-cubic metals. The vector form of this equation is

$$\partial c/\partial t = \text{div}(\mathbf{D} \text{grad } c) = \nabla \cdot (\mathbf{D} \nabla c), \tag{40.15}$$

where  $\mathbf{D}$  has a constant value for cubic crystals, and is a tensor with principal values  $D_1, D_2, D_3$  for anisotropic crystals.

In particular diffusion problems, eqn. (40.14) has to be solved for given initial and boundary conditions. The equation cannot be integrated directly unless the diffusion coefficient is independent of position (i.e. of composition), but many solutions have been obtained

subject to this restriction. Although some of these solutions are required in connection with the theory of diffusional growth described in Sections 54 and 55, we shall give here only an outline treatment of the one-dimensional problems which correspond to common experimental arrangements; for more detailed discussion, reference should be made to specialist texts such as Carslaw and Jaeger (1947) and Crank (1956).

In one dimension, the partial differential equation

$$(\partial c / \partial t) = D(\partial^2 c / \partial x^2) \quad (40.16)$$

has to be solved with given initial and boundary conditions, and the most appropriate form of solution is very dependent on these conditions. In a few experiments, such as the diffusion of a gas through a solid membrane or a thin-walled tube, the flux is constant so that the left-hand side of eqn. (40.16) is zero and the appropriate solution may be obtained directly by integration of (40.13). More general methods of solution include the use of Laplace transforms, the Boltzmann substitution  $\lambda = x/t^{1/2}$ , and the method of separation of variables, but we examine first the use of source functions. If a given component is initially present only as a very thin layer of mass (or number of atoms)  $M$  per unit area at one end of a long specimen, the concentration over a plane at a distance  $x$  from this end after time  $t$  is given by

$$c(x, t) = \{M/(\pi Dt)^{1/2}\} \exp(-x^2/4Dt), \quad (40.17)$$

whereas a thin sandwich of solute in the centre of a long specimen ( $x = 0$ ) will give a distribution of half that of (40.16) in both positive and negative  $x$  directions. This result, which may be verified by inspection, can be used to solve problems in which the initial condition is a uniform or non-uniform extended distribution of solute. Thus a common experimental condition is to have two pure components meeting at an interface, so that the initial conditions are  $c = c_0$  ( $x < 0$ ) and  $c = 0$  ( $x > 0$ ). If the length of the bar is much greater than the mean diffusion distance from the original interface, the boundary conditions are  $c(-\infty, t) = c_0$ ,  $c(+\infty, t) = 0$ . By regarding the initial distribution as a series of infinitesimal planar sources of strengths  $M = c_0 dx'$  situated at distances  $\xi = (x - x')$  from the plane  $x$  where  $0 \geq x' \geq -\infty$ , it follows from (40.17) that

$$\begin{aligned} c(x, t) &= (c_0^2/4\pi Dt)^{1/2} \int_x^\infty \exp(-\xi^2/4Dt) d\xi \\ &= \frac{1}{2}c_0[1 - \operatorname{erf}\{x/2(Dt)^{1/2}\}], \end{aligned} \quad (40.18)$$

where the error function is defined by

$$\operatorname{erf} z = (2/\pi)^{1/2} \int_0^z \exp(-\eta^2) d\eta. \quad (40.19)$$

The distribution may be written more compactly as

$$c(x, t) = \frac{1}{2}c_0 \operatorname{erfc}\{x/2(Dt)^{1/2}\}, \quad (40.20)$$

where the complementary error function is

$$\operatorname{erfc} z = 1 - \operatorname{erf} z = (2/\pi)^{1/2} \int_z^\infty \exp(-\eta^2) d\eta. \tag{40.21}$$

Similarly, if we have an initial distribution of a component given by  $c(-x, 0) = c_0$ ,  $c(+x, 0) = c_1$ , the final distribution becomes

$$\begin{aligned} c(x, t) &= \frac{1}{2}(c_0 + c_1) + \frac{1}{2}(c_1 - c_0) \operatorname{erf}\{x/2(Dt)^{1/2}\} \\ &= c_1 + \frac{1}{2}(c_0 - c_1) \operatorname{erfc}\{x/2(Dt)^{1/2}\}. \end{aligned} \tag{40.22}$$

A generalization of this solution is relevant to a growth problem to be discussed in Section 54. We note that in all cases, the quantity  $(Dt)^{1/2}$ , which has the dimensions of a length, is a measure of the distance travelled by an average diffusing atom, that is of the penetration depth. This is true for all geometrical situations, including initial point and line distributions. The penetration depth, defined as the distance at which the concentration of the diffusing component has decreased to a value of  $1/e$  of its maximum value, is given by  $2(Dt)^{1/2}$ . It is frequently useful to remember this in order of magnitude calculations of diffusion effects.

The substitution  $\lambda = x/t^{1/2}$  first used by Boltzmann leads directly to eqn. (40.18). Thus if  $c(x, t) = f(\lambda)$ , the diffusion equation (40.16) reduces to an ordinary differential equation with solution

$$f(\lambda) = C \int \exp(-\lambda^2/4D) d\lambda. \tag{40.23}$$

The solution is valid provided the boundary conditions can also be expressed in terms of  $\lambda$  alone; if one condition is given for  $\lambda = 0$

$$f(\lambda) - f(0) = 2CD^{1/2} \int_0^{x/2(Dt)^{1/2}} \exp(-\eta^2) d\eta, \tag{40.24}$$

where  $\eta = \lambda/2D^{1/2}$  is a dimensionless integration variable. In order to fix  $C$ , another condition is needed, for example for  $\lambda = \pm \infty$ . Thus the boundary conditions  $f(0) = \frac{1}{2}(c_0 + c_1)$ ,  $f(-\infty) = c_0$ ,  $f(\infty) = c_1$  gives (40.22) from (40.24). The Boltzmann substitution does not apply to the plane source problem (note  $c(x, t)$  in eqn. (40.18)  $\neq f(\lambda)$ ), but it is in general applicable for diffusion problems in infinite or semi-infinite media with uniform or zero initial concentrations. The solutions are usually in the form of error functions or related functions.

When the extent of  $x$  is comparable with the diffusion distance  $(Dt)^{1/2}$ , the above methods of solution become difficult to apply. Whilst it is possible to adapt infinite media solutions by considering repeated reflections at physical boundaries, it is generally preferable to try the method of separation of variables. Thus, if  $c(x, t)$  is written as the product of two functions  $F(x)$  and  $G(t)$ , eqn. (40.16) has a particular solution

$$c(x, t) = (A \sin kx + B \cos Kx) \exp(-k^2Dt)$$

and a general solution

$$c(x, t) = \sum_{n=1}^\infty (A_n \sin k_n x + B_n \cos k_n x) \exp(-k_n^2 Dt). \tag{40.25}$$

This solution is most useful when only the first few terms, preferably only the first, need be retained, and this condition is usually satisfied for times  $t \gtrsim x_t^2/D$ , where  $x_t$  is the total range of  $x$ .

Since  $\operatorname{erf} x = -\operatorname{erf}(-x)$ , the concentration contours given by eqn. (40.18) should be symmetrical about the origin  $x = 0$ . Although this is sometimes approximately true, there is usually marked asymmetry. This may be handled in the formal theory by defining a concentration dependent diffusion coefficient, and using the Boltzmann substitution. The procedure has been especially developed by Matano (1933), who showed how to evaluate a concentration dependent diffusion coefficient from the equation

$$D(c_i) = \frac{1}{2t} \left( \frac{dx}{dc} \right)_{c_i} \int_{c_i}^{c_0} x dc. \quad (40.26)$$

All  $A$  atoms which have left the region  $x < 0$  must have crossed the  $x = 0$  interface and be contained in the region  $x > 0$ . Similarly, all  $B$  atoms which have left the original region  $x > 0$  must be contained in the region  $x < 0$ . Thus we have the condition

$$\int_{c_0}^0 x dc = 0, \quad (40.27)$$

and this may be used to define the position of the  $x = 0$  interface. The meaning of the equation is illustrated in Fig. 9.2, which shows the original and final concentration contours

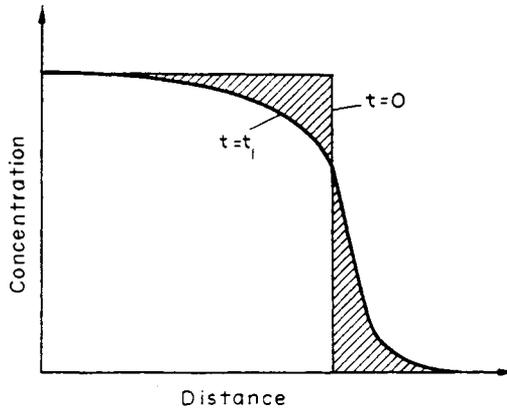


FIG. 9.2. Concentration contours in a typical diffusion experiment. The Matano interface is defined by the condition that the two shaded areas be equal.

of the  $A$  component. The  $x = 0$  interface is defined by the condition that the two shaded areas be equal; it is usually called the Matano interface.

The existence of a Kirkendall effect means that if we plot the concentration contours of  $B$  instead of  $A$ , we shall again find that the shaded areas representing the changes in concentration are equal on each side of the Matano interface, but these areas are not the same as

those of the  $A$  plot. Thus if inert markers are inserted at the original interface, these will move during the experiment relative to the Matano interface, or to the ends of the specimen. As we shall see in the next section, this situation may be described by assigning separate diffusion coefficients to the two components, the movement of the marker interface relative to the Matano interface then being proportional to the difference in these diffusion coefficients. In diffusion experiments of this kind, care should be used about the meaning given by different authors to the term "original" interface. Both the marker interface and the Matano interface may be regarded (in different senses) as the original interface. We should also note that if the diffusion couple consists of two alloys, the Matano interface is still defined by an equation of type (40.27), the limits of integration being from the initial concentration of a component on one side of the interface to its initial concentration on the other.

#### 41. PHENOMENOLOGICAL THEORY OF DIFFUSION

In the original formulation of Fick's law it is assumed that diffusion takes place down the concentration gradient. This is certainly true if the equilibrium state of the assembly is single phase, since at equilibrium corresponding small volumes will contain equal numbers of atoms of a given kind. In any assembly which is not thermodynamically stable, however, the net flow of atoms will result in a lowering of the free energy, and it therefore seems more logical to write the diffusion equations in terms of free energies rather than compositions.

An assembly is in equilibrium with respect to a component  $i$  if the chemical potential  $g_i$  is the same for all parts of all phases of the assembly, and when this condition is attained there will be no net flow of  $i$  atoms. The natural extension of the statistical argument of the last section is thus to assume that diffusion occurs down the chemical potential (or free energy) gradient, rather than down the concentration gradient. This net flow is still the average effect of individual migrations, but these are so weighted that there is a greater probability of motion in the direction which evens out the gradient of free energy.

Darken (1948) developed a phenomenological theory of diffusion based on the above ideas and which included a treatment of the Kirkendall effect. A more general phenomenological theory has been given by Bardeen and Herring (1951, 1952), and has the advantage of being readily applied to an assembly of any number of components. Their treatment, which we shall follow in this section, includes Darken's equations as a special case. A still more general treatment due to Cahn (1961) takes account of additional energy terms due to inhomogeneities in composition, and is discussed in Section 42.

We consider an alloy containing any number of component atoms,  $A, B, C, \dots$ , with  $N$  equivalent sites, of which  $N_i$  are occupied by atoms of the  $i$ th component, and  $N_{\square}$  are vacant. The phenomenological equations are obtained in effect by treating the vacancies as one of the components of the alloy. To minimize the possibility of confusion between our notation  $x_A, x_B, x_C, \dots$ , for atomic fractions and the Cartesian coordinate  $x$ , we shall write most of our equations in this and the following section in three-dimensional vector form, so that  $x$  does not appear explicitly; moreover, we shall use concentrations  $c_A, c_B, \dots$ , instead of atomic fractions wherever possible, and we shall not use the previous abbre-

viation  $x = x_B$ . Care is necessary if the vector equations are transposed into Cartesian form, but this difficulty is not encountered in other chapters.

Our assumption that the diffusion rate is proportional to the gradient of the free energy might lead us to write as the general expression for the diffusion current of the  $i$ th component in place of eqn. (40.2)

$$\mathbf{I}_i = -c_i M'_i \nabla g_i \quad (41.1)$$

where  $g_i$  is the chemical potential per atom and  $c_i$  is the number of  $i$  atoms per unit volume.  $M'_i$  is termed the mobility of  $i$ . The equation thus gives the diffusion current in terms of the average velocity of each atom, expressed as the product of a mobility and an average diffusion "force" on the atom,  $\text{grad } g_i$ . However, this is not the most general expression for the diffusion current. The flux of  $i$  atoms, although mainly dependent on the gradient of the chemical potential  $g_i$ , may also be influenced by the gradients of the chemical potentials of the other components, and of the vacancies if these are present. This concept is in accordance with the general theory of Chapter 4, and we must therefore write the general equation in the form

$$\mathbf{I}_i = -c_j M'_{ij} \nabla g_j - c_{\square} M'_{i\square} / \nabla g_{\square}, \quad (41.2)$$

the first term being summed over all values of  $j$ . In this equation, the most important term is  $M'_{ii}$ , corresponding to  $M'_i$  of eqn. (41.1). The current  $\mathbf{I}_i$  has units  $\text{m}^{-2} \text{s}^{-1}$ , and it is therefore convenient to introduce the mobilities per unit volume,  $M_{ij}$ , which are related to the mobilities per atom  $M'_{ij}$  by the equations

$$M_{ij} = c_{(j)} M'_{i(j)}, \quad (41.3)$$

where the brackets on the subscripts mean, as usual, the suspension of the summation convention. Equation (41.2) can thus be written

$$\mathbf{I}_i = -M_{ij} \nabla g_j - M_{i\square} \nabla g_{\square}, \quad (41.4)$$

In addition to the  $i$  equations giving the net flow of the various atoms in the  $x$  direction, there is obviously an identical equation giving the net vacancy flow  $\mathbf{I}_{\square}$ . The total number of sites will be assumed to remain constant, leading to a condition

$$\mathbf{I}_{\square} = -\sum_i \mathbf{I}_i$$

or

$$M_{\square j} = -\sum_i M_{ij}. \quad (41.5)$$

Equations (41.4) are now of the same form as the general equations (14.3), and it has generally been assumed that the Onsager relations  $M_{ij} = M_{ji}$  are valid. As explained in Chapter 4, this assumption has now been justified by the work of Kirkaldy (1985) and Kirkaldy and Young (1987). It is interesting to note that the reciprocal relations were first applied to the corresponding phenomenological equations for diffusion in electrolytes (**Onsager** and Fuoss, 1932).

The theory presented above does not include the possibility of a marked association between one kind of atom and a vacancy. Such an association is sometimes called a "Johnson molecule" (see p. 404) and may be very important for diffusion in dilute solutions. The statistics of this kind of diffusion are very complex, and it is not at all obvious how the Onsager relations might be applied.

From eqn. (41.5) we obtain  $M_{i\Box} = -\sum_i M_{ij}$ , and eqn. (41.4) can be written in the form

$$\mathbf{I}_i = -M_{ij} \nabla(g_j - g_{\Box}). \quad (41.6)$$

This is the most convenient form for the diffusion equations in a multi-component assembly when vacancies are also present. By making certain assumptions, many more familiar but less general equations may be derived from (41.6). If the vacancies are everywhere maintained in local thermal equilibrium with the structure  $g_{\Box} = 0$  and

$$\mathbf{I}_i = -M_{ij} \nabla g_j. \quad (41.7)$$

Let us consider diffusion in a binary alloy containing *A* and *B* atoms. The atomic fluxes are given by

$$\left. \begin{aligned} \mathbf{I}_A &= -M_{AA} \nabla g_A - M_{AB} \nabla g_B, \\ \mathbf{I}_B &= -M_{AB} \nabla g_A - M_{BB} \nabla g_B. \end{aligned} \right\} \quad (41.8)$$

Now from the Gibbs-Duhem equation,

$$(\partial g_A / \partial \ln c_A) = (\partial g_B / \partial \ln c_B). \quad (41.9)$$

This result depends on the assumption that  $g_{\Box} = 0$ , but will be approximately valid even if this is not true, since the concentration of vacancies is so small. Writing

$$\nabla g_A = (\partial g_A / \partial \ln c_A) (1/c_A) \nabla c_A$$

we have, from (41.8) and (41.9),

$$\begin{aligned} \mathbf{I}_A &= -\frac{\partial g_A}{\partial \ln c_A} \left\{ \frac{M_{AA}}{c_A} - \frac{M_{AB}}{c_B} \right\} \nabla c_A \\ &= -\frac{\partial g_A}{\partial \ln c_A} (M'_{AA} - M'_{AB}) \nabla c_A \\ &= -D_A \nabla c_A, \end{aligned} \quad (41.10)$$

where the intrinsic diffusion coefficient is given by

$$D_A = (\partial g_A / \partial \ln c_A) (M'_{AA} - M'_{AB}) \quad (41.11)$$

and, similarly,

$$D_B = (\partial g_B / \partial \ln c_B) (M'_{BB} - M'_{BA}).$$

Diffusion in a binary assembly may thus be described by two intrinsic diffusion coefficients  $D_A$  and  $D_B$  which are in general unequal, so that the components diffuse at different rates.

The difference between the components arises solely from the mobility terms, and the diffusion force on each atom is equal and opposite. For ternary and higher component assemblies, the Gibbs–Duhem relation does not simplify the equations in this way, so that the average diffusion force varies as well as the mobility. It is not then possible to specify the whole diffusion by means of a single diffusion coefficient for each component.

There will be a mass flow of atoms relative to the lattice, given by the negative of the vacancy current. This flow may be written

$$\mathbf{I}_A + \mathbf{I}_B = -(D_A - D_B) \nabla c_A \quad (41.12)$$

for a binary assembly. Instead of taking the lattice as our reference frame we may use a frame relative to which there is no mass flow. The flux of  $A$  atoms is then

$$\begin{aligned} \mathbf{I}'_A &= \mathbf{I}_A - x_A(\mathbf{I}_A + \mathbf{I}_B) = x_B \mathbf{I}_A - x_A \mathbf{I}_B \\ &= -(x_B D_A + x_A D_B) \nabla c_A. \end{aligned}$$

The quantity

$$D_{\text{chem}} = x_B D_A + x_A D_B \quad (41.13)$$

is the chemical diffusion coefficient as determined in ordinary experiments (e.g. eqn. (40.26)). Instead of specifying the flow by two diffusion coefficients (eqn. (41.11)), we can therefore use a single coefficient and a velocity of mass flow eqns. ((41.12) and (41.13)). The two methods are equivalent, although the first seems generally preferable. Equations (41.11)–(41.13) form Darken's phenomenological equations for binary diffusion when there is a Kirkendall effect; they depend only on the assumption that the vacancies are maintained in local thermal equilibrium. The validity of the assumption is discussed in Section 43. Seitz has shown that the equations actually depend on the vacancy concentration being a function only of the composition.

As stated on p. 382, the Kirkendall effect provides an experimental test to distinguish between place exchange mechanisms and vacancy or interstitial mechanisms. If diffusion is by some form of place exchange, then in place of eqn. (41.5) we have the condition  $\sum_i \mathbf{I}_i = 0$ . There is then no net flow of matter relative to the lattice, and

$$D_A = D_B = D_{\text{chem}}.$$

Darken also derived relations between the intrinsic diffusion coefficients and the tracer diffusion coefficients. Suppose we have an alloy in which the concentration of  $A$  and  $B$  atoms is uniform but in which the  $A$  atoms consists of two isotopes  $A$  and  $A_+$  such that there is a concentration gradient. The currents of  $A$  and  $A_+$  atoms will be given by (41.10) as

$$\left. \begin{aligned} \mathbf{I}_A &= -(\partial g_A / \partial \ln c_A) (M'_{AA} - M'_{AA_+}) \nabla c_A \\ \mathbf{I}_{A_+} &= -(\partial g_{A_+} / \partial \ln c_{A_+}) (M'_{A_+A_+} - M'_{A_+A}) \nabla c_{A_+} \end{aligned} \right\} \quad (41.14)$$

Since the isotopes are chemically equivalent, the chemical potentials depend only on the entropy of mixing, and

$$\begin{aligned} g_A &= kT \ln x_A = kT \ln(c_A/c), & g_{A_+} &= kT \ln(c_{A_+}/c), \\ (\partial g_A / \partial \ln c_A) &= (\partial g_{A_+} / \partial \ln c_{A_+}) = kT. \end{aligned}$$

The diffusion currents  $I_A, I_{A_+}$  must be equal and opposite. If we neglect the cross-terms, we thus have  $M'_{AA} = M'_{A_+A_+}$ , and the diffusion coefficient may be written

$$D_{A_+} = M'_{AA}kT. \tag{41.15}$$

This relation between mobility and diffusion coefficient is known as Einstein's equation, and was first derived in 1905. It is valid for any component of an ideal solution provided the cross-terms can be neglected.

We also have from eqns. (41.11) and (41.15)

$$\frac{D_A}{D_{A_+}} = \frac{\partial(g_A/kT)}{\partial \ln c_A} = \frac{D_B}{D_{B_+}}. \tag{41.16}$$

This is Darken's thermodynamic relation, and also depends on the approximation of neglecting  $M_{AA_+}, M_{AB}$ , etc. The quantities  $D_A, D_{A_+}$  are respectively the intrinsic coefficient and the tracer diffusion coefficient of  $A$  in an alloy of  $A$  and  $B$ .

The extent of the approximation involved in neglecting the cross-terms was examined by Bardeen and Herring in the following way. Consider the simple case of radioactive diffusion in pure  $A$ . Equations (41.14) are valid, and we may write the diffusion coefficients

$$\begin{aligned} D_A &= kT(M'_{AA} - M'_{AA_+}) = D_{AA} - D_{AA_+}, \\ D_{A_+} &= kT(M'_{A_+A_+} - M'_{A_+A}) = D_{A_+A_+} - D_{A_+A}. \end{aligned}$$

We also know that

$$M'_{AA_+} = M_{AA_+}/c_{A_+}, \quad M'_{A_+A} = M_{AA_+}/c_A,$$

and if the tracer  $A_+$  is present in very small concentration we may thus neglect  $M'_{A_+A}$  in comparison with  $M'_{AA_+}$ . Since  $D_A = D_{A_+}$ , we therefore have

$$D_{A_+} = D_{A_+A_+} = D_{AA} - D_{AA_+}. \tag{41.17}$$

Now suppose that we have pure  $A$  with a concentration gradient of vacancies, so that  $g_{\square} = kT \ln(c_{\square}/\bar{c}_{\square})$ , where  $\bar{c}_{\square}$  is the equilibrium vacancy concentration. The vacancy current given by eqn. (41.4) is thus

$$I_{\square} = -I_A = M_{AA} \nabla g_{\square} = (kTM_{AA}/c_{\square}) \nabla c_{\square}.$$

The diffusion coefficient for vacancies is thus

$$D_{\square} = kTM_{AA}/c_{\square} = (c_A/c_{\square}) kTM_{AA} = D_{AA}/x_{\square},$$

since  $c_{\square}$  is small. Comparing this equation with (41.17),

$$D_{A_+} = x_{\square}D_{\square} - D_{AA_+}. \tag{41.18}$$

Now a simple kinetic argument has already been given on p. 385 to deduce  $D_{A_+} = x_{\square}D_{\square}$ . If the motion of a tracer atom can be treated as a random walk, the probability per unit time that it will move to a neighbouring site is equal to the frequency with which a vacancy jumps, multiplied by the probability  $x_{\square}$  that the site is vacant. The diffusion coefficients

should thus be in this ratio. The defect in the reasoning lies in the assumption which we explicitly made on p. 386 that the successive jumps are uncorrelated. If an atom has just left a given site, there will be some time during which the probability of a reverse jump is greater than that of a further jump to another site. This is because the probability of the original site remaining vacant is greater than that of one of the remaining sites becoming vacant. This correlation results in  $D_{A+}$  being less than  $x_{\square}D_{\square}$ , and the difference is just the term  $D_{AA+}$ . For vacancy self-diffusion, we thus see that the correct physical interpretation of the vacancy cross-terms is the effect of one atomic movement on the probabilities of further movements.

Le Claire and Lidiard (1956) considered the correlation effects for various kinds of diffusion mechanism, using eqn. (40.7). For a vacancy mechanism, there is *direct* correlation only between successive jumps, i.e. the probability of a given jump being in any particular direction depends only on the direction of the jump immediately preceding the one considered. This means that

$$\overline{\cos \theta_j} = \overline{\cos \theta_{j-1}} \overline{\cos \theta_1} = (\overline{\cos \theta_1})^j. \quad (41.19)$$

Substituting into eqn. (40.7) and allowing  $n$  and  $t$  to become large then gives

$$\overline{S^2} = nr_1^2 f, \quad \text{or} \quad D = \frac{1}{6} \Gamma r_1^2 f \quad (41.20)$$

where the "correlation factor"

$$f = (1 + \overline{\cos \theta_1}) / (1 - \overline{\cos \theta_1}). \quad (41.21)$$

Note that the magnitude of the cross-terms is given by  $M'_{AA+} = 1 - f$ .

A rough estimate of  $\overline{\cos \theta_1}$  is obtained by considering only the first jump of a vacancy with which a moving atom has just changed places. For a random walk of the vacancy, the probability is  $1/z$  that it will re-exchange places with the same atom, and since  $\theta_1 = -\pi$  for this jump  $\overline{\cos \theta_1} = -1/z$  and

$$f \simeq 1 - 2/z. \quad (41.22)$$

For a more accurate computation, we let  $P_k$  be the probability that the next jump of the tracer atom will be along the direction  $\theta_{(1)k}$  to the  $k$ th site around the atom. We must consider all possible paths which transfer the vacancy from site 1 to site  $k$ , and, in general,

$$P_k = \sum_{i=1}^{\infty} n_{ik} z^{-i}, \quad (41.23)$$

where  $n_{ik}$  is the number of paths along which the  $(i-1)$ th jump of the vacancy takes it to site  $k$  and  $z^{-i}$  is the cumulative probability of the vacancy moving along each such path and eventually changing places with the tracer atom. Then

$$\overline{\cos \theta_1} = \sum_{k=1}^z P_k \cos \theta_{(1)k}. \quad (41.24)$$

Equation (41.23) is based on the explicit assumption that the vacancy executes a random walk. This is not strictly valid even for tracer self-diffusion if the jump frequency of the isotope used as a tracer is not identical with that of the abundant isotope. Since the chemical properties are identical, changes in the jump frequency arise only because different isotopes have different masses. The correction to  $f$  from this *isotope effect* is very small, but may nevertheless be used to determine  $f$  experimentally (see below).

The difficulty in evaluating  $\cos \theta_1$  from eqn. (41.24) is that the series for  $P_k$  converges rather slowly. Bardeen and Herring used a numerical technique and Le Claire and Lidiard (1956) developed a matrix method which has also been discussed by Mullen (1961) and Manning (1968). An alternative method, based on an electrical analogue, was used by Compaan and Haven (1958) who give values of 0.5, 0.653, 0.727, and 0.781 for the correlation factor  $f$  in diamond cubic, simple cubic, b.c.c., and f.c.c. crystals respectively. Equation (41.22) is thus exact for the diamond structure but gives values which are slightly too high for the other structures.

Correlation effects do not occur for diffusion by the ring mechanism and simple interstitial mechanisms, but they are expected for interstitialcy diffusion. For self-diffusion in cubic crystals  $f$  is a numerical factor which, apart from the small isotope effect, depends only on the diffusion mechanism and crystal structure; its importance lies not so much in the magnitude of the correction but in the realization that eqn. (41.1) is not strictly valid even for one of the simplest diffusion problems. Moreover, measurements of the correlation factor may enable the diffusion mechanism to be deduced.

Equation (41.19) cannot be used for self-diffusion in non-cubic structures and for divacancy diffusion in any structure. The evaluation of the correlation factor is then much more difficult, but has been accomplished in several cases of interest by Mullen (1961), Howard (1966), Bakker (1971), and Mehrer (1972). Mehrer obtains  $f = 0.468$  for divacancy self-diffusion in f.c.c. metals. It is important to realize that  $f$  is not a purely geometrical factor in non-cubic crystals, since transitions with different jump frequencies are then possible. Expressions for the correlation factors for diffusion in different directions depend on ratios of jump frequencies, and may be temperature-dependent.

Correlation effects in chemical diffusion are often much more significant than in self-diffusion since the approximation of regarding the motion of a vacancy as a random walk is then invalid. We consider this further on p. 404.

## 42. UPHILL DIFFUSION

We shall now examine a little more fully the form of eqn. (41.11), giving the diffusion coefficient in a binary alloy. Since  $M'_{AA}$  is necessarily greater than  $M'_{AB}$ , the latter being neglected in first approximation, the term in brackets is necessarily positive. The condition for "uphill" diffusion, i.e. for  $D_A$  to be negative, is therefore

$$(\partial g_A / \partial c_A) < 0, \quad \text{i.e.} \quad (\partial g_A / \partial x_A) < 0. \quad (42.1)$$

We have imposed no restrictions on the independent variables determining  $g_A$ , which may be a function not only of temperature and concentration but also of other external

constraints and internal stress fields. One possible source of internal stress is the variation of lattice parameter with composition, and if this is included in the expression for the chemical potential, the inequality (42.1) gives a very general condition for uphill diffusion, which is valid for any imposed conditions of the assembly.

The inequality (42.1) governs uphill diffusion of one component, but it is generally more useful to consider uphill chemical diffusion, i.e. spontaneous segregation of two components, the condition for which is a negative value of  $D_{\text{chem}}$ . Applying eqn. (41.12) to the flux of  $B$  atoms, and neglecting the off diagonal terms in eqn. (41.8),

$$\begin{aligned} \mathbf{I}'_B &= \mathbf{I}_B - x_B(\mathbf{I}_A + \mathbf{I}_B) \\ &= -c_B x_A (M'_{BB} \nabla g_B - M'_{AA} \nabla g_A) \end{aligned} \quad (42.2)$$

since  $c_B x_A = c_A x_B = x_A x_B / v$ , where  $v$  is the atomic volume. This equation may also be written in the form

$$\mathbf{I}'_B = -M_{\text{chem}} \nabla (g_B - g_A) = -M_{\text{chem}} \nabla \varphi_D, \quad (42.3)$$

where  $\varphi_D$  may be regarded as the "driving force" for diffusion and

$$\begin{aligned} M_{\text{chem}} &= c_B x_A \{x_A M'_{BB} + x_B M'_{AA}\} \\ &= x_A^2 M_{BB} + x_B^2 M_{AA} \end{aligned} \quad (42.4)$$

is a mobility per unit volume; the remaining part of the right-hand side of (42.2) vanishes because of the Gibbs-Duhem equation. Now since  $g_B - g_A = \partial g / \partial x_B$  (eqn. (22.7)), it follows that

$$D_{\text{chem}} = M_{\text{chem}} v (\partial^2 g / \partial x_B^2), \quad (42.5)$$

and for uphill diffusion,

$$\partial^2 g / \partial x_B^2 = v^{-2} (\partial^2 g / \partial c^2) < 0, \quad (42.6)$$

which means that the temperature and composition must lie inside the chemical spinodal (eqn. (22.42)). For condensed phases, (42.6) may be replaced by

$$\partial^2 f / \partial c^2 < 0, \quad (42.7)$$

where  $f$  is the Helmholtz free energy per atom. In the following development we use  $c$  rather than  $x_B$  as the composition variable (see p. 391), even though this introduces more terms involving the atomic volume than would otherwise be necessary.

Uphill diffusion inside the spinodal was discussed by Becker (1937) and later workers, but the inadequacy of the diffusion equation (42.3) was not realized until the work of Cahn (1961, 1962). Difficulties arise because composition gradients are established over fairly short distances during spinodal decomposition and the gradient and coherency energy terms discussed in Section 22 can then not be neglected. In the isotropic approximation, the free energy per atom of an inhomogeneous solid solution is given by

$$g_{ih} = V^{-1} \int [g(c) + v^3 \{ \varepsilon^2 Y'(c - c_0)^2 + K(\nabla c)^2 \}] dV, \quad (42.8)$$

where the integral is taken over the volume of the assembly and  $\varepsilon$ ,  $Y'$ , and  $K$  have the meanings defined on p. 185. For equilibrium of the non-homogeneous assembly this integral must be minimized subject to the requirement

$$\int (c - c_0) dV = 0, \tag{42.9}$$

which ensures that the average concentration  $c_0$  is constant. The functional minimization by Euler's equation leads to the definition of a potential

$$\varphi_D = v^{-1}(\partial g / \partial c) + v^2\{2\varepsilon^2 Y' (c - c_0) - 2K \nabla^2 c - (\partial K / \partial c) (\nabla c)^2\}, \tag{42.10}$$

which has a constant value at equilibrium. The first term in  $\varphi_D'$  is  $\varphi_D$  of eqn. (42.3), and we may now generalize that equation by assuming the diffusion flux to be proportional to the gradient of  $\varphi_D'$  instead of  $\varphi_D$ . This gives

$$I_D = -M_{\text{chem}}\{v^{-1}(\partial^2 g / \partial c^2) + 2\varepsilon^2 Y' v^2\} \nabla c + 2M_{\text{chem}} K v^2 \nabla^3 c + M_{\text{chem}} v^2 \nabla\{(\partial K / \partial c) (\nabla c)^2\} \tag{42.11}$$

and

$$\begin{aligned} (\partial c / \partial t) = & M_{\text{chem}}\{v^{-1}(\partial^2 g / \partial c^2) + 2\varepsilon^2 Y' v^2\} \nabla^2 c - 2M_{\text{chem}} K v^2 \nabla^4 c \\ & + \left[ M_{\text{chem}} \frac{\partial}{\partial c} \{v^{-1}(\partial^2 g / \partial c^2) + 2\varepsilon Y' v^2\} (\nabla c)^2 + \dots \right]. \end{aligned} \tag{42.12}$$

The last term of eqn. (42.11) and the non-linear terms in square brackets of eqn. (42.12) may be neglected if  $c - c_0$  is small. The two equations are then linear but differ from the classical forms of Fick's first and second laws because of the terms in  $\nabla^3 c$  and  $\nabla^4 c$  respectively.

A particular solution of the linear form of (42.12) may be written as a sine wave, and the general solution is obtained by superposition of all such solutions. This gives

$$c(\mathbf{r}, t) - c_0 = \iiint A(\boldsymbol{\beta}, t) \exp(-i\boldsymbol{\beta} \cdot \mathbf{r}) d\boldsymbol{\beta}, \tag{42.13}$$

where  $\mathbf{r}$  is the position vector,  $\boldsymbol{\beta}$  is a wave vector representing a Fourier component of the composition variation with wavelength  $\lambda = 2\pi/\beta$ , and  $A(\boldsymbol{\beta}, t)$  is the amplitude of this component at time  $t$ . By substituting back into the linearized diffusion equation, we find

$$A(\boldsymbol{\beta}, t) = A(\boldsymbol{\beta}, 0) \exp\{R(\boldsymbol{\beta})t\}, \tag{42.14}$$

where the initial fluctuation is specified by the amplitude  $A(\boldsymbol{\beta}, 0)$  of the various components and  $R(\boldsymbol{\beta})$  is called the amplification factor and is given by

$$R(\boldsymbol{\beta}) = -M_{\text{chem}}\beta^2\{v^{-1}(\partial^2 g / \partial c^2) + 2\varepsilon^2 Y' v^2 + 2K v^2 \beta^2\}. \tag{42.15}$$

In this equation  $M$  is inherently positive and  $K$  is expected to be positive for a system with a spinodal; negative  $K$  occurs for ordering systems but cannot be handled by means of a continuum theory. Thus the condition for a particular Fourier component to increase in

amplitude with time is that  $R(\beta)$  is positive. This is only possible if the mean composition lines inside the curve

$$(\partial^2 g / \partial c^2) = -2\epsilon^2 Y' v^3 \quad (42.16)$$

and if the wave vector has amplitude

$$\beta^2 < \beta_c^2 = -\{v^{-3}(\partial^2 g / \partial c^2) + 2\epsilon^2 Y'\} / 2K. \quad (42.17)$$

A positive amplification factor implies uphill diffusion, and we see that a necessary condition is that the composition and temperature lie inside the coherent spinodal since eqn. (42.16) is identical with (22.45). The coherency energy in eqn. (42.8) has two effects on diffusion processes; it restricts the region of negative diffusion to the area inside the coherent spinodal instead of that inside the chemical spinodal, and when an anisotropic calculation is made it introduces directionality into diffusion, even in cubic crystals. The gradient energy does not restrict the temperature or composition range of negative diffusion, but ensures that it can only occur over distances greater than  $\lambda_c = 2\pi/\beta_c$ . A typical value for  $\lambda_c$  is 10 nm.

Cahn's generalized diffusion equation may be regarded as a correction for the assumption in Fick's law that the flux is proportional to the concentration gradient. However, the modifications he introduced have negligible effects in most situations where the diffusion distance is of the order of microns, and its application in practice tends to be restricted mainly to discussion of spinodal decomposition. We defer further development of this theory to Part II, Chapter 18, although we note that an atomistic (regular solution) model of spinodal decomposition was provided by Hillert (1961) prior to the work of Cahn. Hillert's development includes the gradient energy but not the coherency energy.

Despite the limitation mentioned above, it has proved possible to test the new diffusion equation in non-spinodal situations by preparing multi-layered sandwiches of two components by vapour deposition. Individual layers of 1–3 nm may be deposited successively, and the homogenization process on annealing can then be followed by suitable X-ray techniques. In work on gold–silver alloys (Cook and Hilliard, 1969) and copper–palladium (Philofsky and Hilliard, 1969) confirmation of both gradient energy and coherency energy effects was obtained and values of the gradient energy coefficient  $K$  were measured.

### 43. KINETIC THEORY OF VACANCY DIFFUSION

The elementary kinetic theory of Section 40 led to Fick's law; we shall now replace it by a more satisfactory theory which leads to the phenomenological equations of Section 41. This account is due initially to Seitz (1948, 1950a), and is important in showing the physical significance of the terms in the formal theory. We restrict the discussion to vacancy diffusion, although the method can readily be applied to the other possible mechanisms, and we again consider one-dimensional flow in a binary alloy.

The numbers of  $A$  atoms,  $B$  atoms, and vacancies on two neighbouring planes, distance  $d$  apart, may be written  $c_{A1}d$ ,  $c_{B1}d$ ,  $c_{\square 1}d$ ,  $c_{A2}d$ ,  $c_{B2}d$ , and  $c_{\square 2}d$  respectively. Equation (40.1) gives the net current of  $A$  atoms, and the vacancy current is similarly

$$I_{\square} = d(c_{\square 1}\pi_{\square 12} - c_{\square 2}\pi_{\square, 21}). \quad (43.1)$$

Since  $c_{\square}$  is small, even though not necessarily constant,  $c_B$  is effectively determined by  $c_A$ . The probability per unit time that a vacancy will jump from a plane where the concentration of  $A$  atoms is  $\zeta$  to a neighbouring plane where it is  $\eta$  is thus entirely determined by  $\zeta$  and  $\eta$ , and may be written  $\pi_{\square}(\zeta\eta)$ . In particular,  $\pi_{\square, 12} \equiv \pi_{\square}(c_{A1}, c_{A2})$ .

When the vacancy jumps from the region  $\zeta$  to  $\eta$ , it may change places with either an  $A$  atom or a  $B$  atom. We write the frequency with which it changes places with an  $A$  atom as  $\pi_{\square A}(\zeta\eta)$ .<sup>†</sup> Clearly

$$\pi_{\square}(\zeta\eta) = \pi_{\square A}(\zeta\eta) + \pi_{\square B}(\zeta\eta). \tag{43.2}$$

In order to avoid confusion, we note that the terms in the bracket give the direction in which the vacancies move, and are opposite to the direction in which the atoms move.

Since  $c_{A1}, c_{A2}$  will be only slightly different on neighbouring planes, the two frequencies in eqn. (43.1) will both be nearly equal to  $\pi_{\square}(c_A c_A)$ , which is the frequency of vacancy jumps between two neighbouring planes both containing  $c_A d = \frac{1}{2}(c_{A1} + c_{A2})d$  atoms of type  $A$  per unit area. We can thus express the difference between  $\pi_{\square}(c_{A1}c_{A2})$  and  $\pi_{\square}(c_{A2}c_{A1})$  by regarding  $\pi_{\square}(\zeta\eta)$  as a continuous function of  $\zeta$  and  $\eta$ , and expanding in a Taylor series about  $\zeta = \eta = c_A$ . This gives

$$\left. \begin{aligned} \pi_{\square}(c_{A1}c_{A2}) &= \pi_{\square}(c_A c_A) - \frac{d}{2} \frac{\partial \pi_{\square}}{\partial \zeta} \frac{\partial c_A}{\partial x} + \frac{d}{2} \frac{\partial \pi_{\square}}{\partial \eta} \frac{\partial c_A}{\partial x}, \\ \pi_{\square}(c_{A2}c_{A1}) &= \pi_{\square}(c_A c_A) + \frac{d}{2} \frac{\partial \pi_{\square}}{\partial \zeta} \frac{\partial c_A}{\partial x} - \frac{d}{2} \frac{\partial \pi_{\square}}{\partial \eta} \frac{\partial c_A}{\partial x}. \end{aligned} \right\} \tag{43.3}$$

The differentials  $\partial \pi_{\square} / \partial \zeta, \partial \pi_{\square} / \partial \eta$  have, of course, to be evaluated at  $\zeta = \eta = c_A$ . Substituting into (43.1), we obtain the vacancy current

$$I_{\square} = d \left[ \pi_{\square}(c_A, c_A) \{c_{\square 1} - c_{\square 2}\} + \frac{d}{2} \{c_{\square 1} + c_{\square 2}\} \left\{ -\frac{\partial \pi_{\square}}{\partial \zeta} + \frac{\partial \pi_{\square}}{\partial \eta} \right\} \frac{\partial c_A}{\partial x} \right]. \tag{43.4}$$

Purely for convenience, we now introduce the notation

$$\frac{\partial \pi_{\square}}{\partial c_A} = -\frac{\partial \pi_{\square}}{\partial \zeta} + \frac{\partial \pi_{\square}}{\partial \eta}. \tag{43.5}$$

Writing also  $c_{\square} = \frac{1}{2}(c_{\square 1} + c_{\square 2})$ ,

$$I_{\square} = -d^2 \pi_{\square}(c_A, c_A) \frac{\partial c_{\square}}{\partial x} + d^2 c_{\square} \frac{\partial \pi_{\square}}{\partial c_A} \frac{\partial c_A}{\partial x}. \tag{43.6}$$

There may thus be a net vacancy current, even when there is no concentration gradient of vacancies.

It should be noted that the term  $\partial \pi_{\square} / \partial c_A$  is defined by (43.5); its use does not imply that  $\pi_{\square}$  is a function of a single variable  $c_A$ . In some papers on diffusion theory, this latter

<sup>†</sup> It should be noted that  $\pi_{\square A}(\zeta\eta)$  is not equal to  $\pi_{A, 21}$ , used in Section 40; the two frequencies are in the ratio  $c_{A2} : c_{\square 1}$ .

assumption has in fact been made, and an equation of the same form as (43.6) is obtained. Seitz's theory is more general, insofar as it is recognized that  $\partial\pi/\partial\zeta$  is not necessarily equal to  $-\partial\pi/\partial\eta$ .

We can now write the current of  $A$  atoms as

$$I_A = d[c_{\square 2}\pi_{\square A}(c_{A2}, c_{A1}) - c_{\square 1}\pi_{\square A}(c_{A1}, c_{A2})].$$

Expanding the  $\pi_{\square A}$  in exactly the same way as we expanded the  $\pi_{\square}$  terms, we obtain

$$I_A = d^2\pi_{\square A}(c_A, c_A) \frac{\partial c_{\square}}{\partial x} - d^2c_{\square} \frac{\partial \pi_{\square A}}{\partial c_A} \frac{\partial c_A}{\partial x}, \quad (43.7)$$

where the term  $\partial\pi_{\square A}/\partial c_A$  is defined by an equation of the type of (43.5). Similarly,

$$I_B = d^2\pi_{\square B}(c_A, c_A) \frac{\partial c_{\square}}{\partial x} - d^2c_{\square} \frac{\partial \pi_{\square B}}{\partial c_A} \frac{\partial c_A}{\partial x}, \quad (43.8)$$

and in view of (43.2)

$$I_A + I_B + I_{\square} = 0, \quad (43.9)$$

which ensures the conservation of lattice sites.

Equations (43.6)–(43.8) correspond to the general phenomenological eqn. (41.6), in which no assumption is made about  $g_{\square}$ .

If we now assume that the concentration of vacancies is determined by the concentration of  $A$  atoms, we may write

$$\frac{\partial c_{\square}}{\partial x} = \frac{\partial c_{\square}}{\partial c_A} \frac{\partial c_A}{\partial x}. \quad (43.10)$$

On substituting into (43.7) and (43.8), we obtain the diffusion coefficients

$$\left. \begin{aligned} D_A &= -d^2 \left[ \pi_{\square A}(c_A, c_A) \frac{\partial c_{\square}}{\partial c_A} + c_{\square} \frac{\partial \pi_{\square A}}{\partial c_A} \right], \\ D_B &= d^2 \left[ \pi_{\square B}(c_A, c_A) \frac{\partial c_{\square}}{\partial c_A} + c_{\square} \frac{\partial \pi_{\square B}}{\partial c_A} \right], \end{aligned} \right\} \quad (43.11)$$

and from (43.6)

$$I_{\square} = -(D_A - D_B) \frac{\partial c_A}{\partial x}. \quad (43.12)$$

We thus obtain Darken's phenomenological equations if (43.10) is correct. This condition is slightly more general than the assumption of thermal equilibrium for the vacancies, used in deriving the same equations in Section 41.

Darken's thermodynamic relations (eqn. (41.16)) were obtained by Bardeen on the assumption that the probability of a vacancy changing places with an  $A$  atom is proportional to the number of  $A$  atoms on the plane to which it jumps. This implies that each plane normal to the diffusion flow is in equilibrium, so that the probability of a given site being occupied by a given kind of atom is equal to the concentration of such atoms in the plane

considered. This equilibrium must be established immediately after each atom has moved. In fact, as we have seen, correlation effects give a significantly higher probability to an immediate return jump, and in self-diffusion, the error in neglecting these terms is of the order  $1/z$ . Correlation effects in chemical diffusion are discussed below.

Plausible arguments were given by Bardeen and Herring (1952) to show that it is probable that the concentration of vacancies can be maintained close to its equilibrium value by the action of edge dislocations. If this is true, Darken's equations are valid, and lattice sites are created on one side of a diffusion couple and destroyed on the other, thus giving a net movement of the marker interface relative to the Matano interface. The mechanism envisaged for the maintenance of the equilibrium concentration of vacancies was the operation of dislocation sources in climb, as described on p. 250. However, there is now a large body of evidence to show that porosity is developed on one side of the original interface in nearly all Kirkendall effect experiments, and this must mean that local equilibrium of vacancy concentrations cannot be fully preserved. This has been demonstrated directly by Heumann and Walther (1957) who analysed the movements of inert markers distributed throughout the diffusion zone.

At first, it seems rather strange that vacancies in diffusion experiments should condense to form macroscopic spherical holes rather than flat discs which collapse into prismatic dislocations as in quenched specimens. The porosity is not a negligible effect; in typical experiments (e.g. Barnes, 1952) about one-half of the net vacancy current has been used in forming voids. Using arguments essentially equivalent to those of the classical theory of nucleation described in the next chapter, Seitz (1953) concluded that homogenous nucleation of pores by aggregation of single vacancies is most improbable, since this requires a relative excess vacancy concentration, defined as  $(x_{\square}/\bar{x}_{\square}) - 1$ , to be  $\sim 100$ . An analysis of the then available experimental data led to the conclusion that relative excess vacancy concentrations of  $\sim 1$  are possible, and give a lifetime of  $\sim 10^{11}$  vacancy jumps. However, if condensation on relatively large defects occurs, it is possible that the maximum excess concentration is much less than this. Resnick and Seigle (1957) have shown that oxide inclusions might act as nuclei for pore formation, and Barnes and Mazey (1958) demonstrated that an external pressure is able to prevent pore formation, whilst a larger pressure will collapse the voids produced by the diffusion process. These workers and others (e.g. Balluffi, 1954) also concluded that pore nucleation is heterogeneous, and that the vacancy supersaturation is of the order of 1%.

The full details of the processes leading to the Kirkendall effect are not yet certain; the whole phenomenon is rather complex, and various secondary reactions such as polygonization in the diffusion zone have been detected. The observed porosity accounts for only a fraction of the excess vacancies, generally about half, so that the previous conclusion that dislocations act as sources and sinks for vacancies in such experiments is still valid, even if the vacancy concentration does depart appreciably from its equilibrium value. There is also some experimental evidence that the climbing of dislocations to the free surface produces observable steps (Barnes, 1955).

In real crystal structures, the Bardeen-Herring climb mechanism needs modification since edge dislocations are not usually the boundaries of single atomic planes. The usual

picture of a f.c.c.  $\frac{1}{2}a\langle 110 \rangle$  edge dislocation, for example, requires the insertion of two extra atomic  $\{110\}$  planes above the  $\{111\}$  glide plane. If the dislocation is undissociated in the glide plane, climb could take place by condensation of vacancies in a double layer, but if the dislocation dissociates into two Shockley partials, the extra atomic planes separate to opposite ends of the stacking fault, and the climb process is more complex. Analysis of results of thin film observations suggests that glide dislocations climb easily only when they are parallel to  $\langle 110 \rangle$ , the vacancies being added as a single layer on the intersecting  $\{111\}$  plane, and also when they are almost parallel to an intersecting  $\{110\}$  plane, the vacancies being added as a double layer. This is contrary to the original concept that an edge or mixed dislocation just climbs normal to its slip plane.

The kinetic theory described in this section expresses the diffusion rate in terms of the vacancy concentration and of the different probabilities that an  $A$  atom and a  $B$  atom will change places with a neighbouring vacancy in unit time, but it does not include correlation effects. Let us now consider a dilute solution of  $B$  in  $A$ . It is found experimentally that tracers of the impurity ( $B$ ) atoms may diffuse at rates which are considerably different from the self-diffusion rate of the solvent atoms, and which may in general be either greater or smaller. Part of the explanation for this effect will arise from the change in the energy of formation of a vacancy in the neighbourhood of an impurity atom, approximate treatments being possible in terms of electrostatic interactions between the impurity and the lattice or in terms of the misfitting sphere model. Another factor of importance is the correlation between the motion of the solute atom and that of the solvent atoms, the rate of jumping of the solute atom being limited either by the activation energy for its own motion, or that for the motion of the solvent atom. As first suggested by Johnson (1941), there may be a strong association between vacancies and impurity atoms. Instead of the random migration of vacancies, assumed in the treatment above, the solute-atom-vacancy pair ("Johnson molecule") will then move together through the lattice, thus increasing the diffusion coefficient. Correlation effects of this kind are of major importance in contrast to the minor corrections to the random walk theory required by correlation effects in the self-diffusion of a pure component.

For self-diffusion in a cubic pure metal, eqn. (40.8) reduces to  $D = a^2 P k_A$ . Le Claire and Lidiard (1956) treated the problem of solute diffusion by the Johnson mechanism in a f.c.c. structure, using the approach outlined on p. 396. We let  $P$  now be the probability of finding a vacancy as the nearest neighbour of a solute  $B$  atom. The rates of interchange of this vacancy with the  $B$  atom, with the four  $A$  atoms which are nearest neighbours to both vacancy and  $B$  atom ( $A_1$  sites) and with the seven  $A$  atoms which are neighbours of the vacancy, but not of the  $B$  atom ( $A_2$  sites) are assumed to be all different, and are denoted by  $k_B$ ,  $k_{A_1}$ , and  $k_{A_2}$  respectively. The mean value of  $\cos \theta_1$  in eqn. (41.20) is found by a method similar to that discussed previously and, the diffusion coefficient is expressed as

$$D_B = a^2 P k_B (k_{A_1} + \frac{7}{2} k_{A_2}) / (k_B + k_{A_1} + \frac{7}{2} k_{A_2}). \quad (43.13)$$

When the rate of jump of the  $B$  atom is much greater than that of the solvent atoms, this expression takes the limiting form

$$D_B = a^2 P (k_{A_1} + \frac{7}{2} k_{A_2}) \quad (43.14)$$

and is independent of  $k_B$ . The diffusion rate of the solute atom is then limited only by the jump frequency of the solvent atoms, but the solute will diffuse faster than the solvent. The other limiting form is when the  $B$  atoms take much longer to jump into neighbouring vacancies than do the  $A$  atoms. Equation (43.13) then has the limiting form

$$D_B = a^2 P k_B \quad (43.15)$$

and the diffusion rate of the solute is generally less than that of the solvent. An interesting reduction to the case of self-diffusion in a pure metal is obtained by writing  $k_B = k_{A1} = k_{A2}$  in eqn. (43.13). This gives

$$D_A = \frac{9}{11} a^2 P k_A, \quad (43.16)$$

the factor (9/11) being the value of the correlation factor eqn. (41.21) previously derived by Bardeen and Herring.

The Le Claire-Lidiard calculation is based on the assumption that a vacancy which jumps to an  $A2$  site is completely dissociated from the solute atom. In a more accurate treatment, Manning (1962, 1964) introduced another local frequency  $k_{A3}$  which describes an association jump (the reverse of the dissociation jump) and he further assumed that all jumps of the vacancy other than those described above take place with the frequency  $k_A$  which applies in pure solvent. He obtained for the correlation factor

$$f_B = (k_{A1} + \frac{7}{2} F k_{A2}) / (k_B + k_{A1} + \frac{7}{2} F k_{A2}), \quad (43.17)$$

where  $F$  is a rather complex function of  $k_{A3}/k_A$  which reduces to  $F = 1$  for  $k_{A2} \ll k_A$ ,  $F \approx 0.74$  for  $k_{A2} = k_A$ , and  $F = \frac{2}{7}$  for  $k_{A2} \gg k_A$ .

Unless the atomic forces are very short-ranged, the assumption that there is a single frequency for vacancy jumps to each of the seven  $A2$  positions is doubtful since these sites are variously second, third, or fourth neighbours of the solute atom. The error caused by this assumption is likely to be much more serious in b.c.c. structures where there are no  $A1$  type sites but where the difference between first and second neighbour distances is not great (see p. 120). If the vacancy jump to a site which is the second neighbour of the solute atom has the same frequency as the jump to any other  $A2$ -type site, the correlation factor for solute diffusion in b.c.c. structures is

$$f_B = 7 F k_{A2} / (2 k_B + 7 F k_{A2}), \quad (43.18)$$

where  $F$  is again a function of  $k_{A3}/k_A$ . This expression becomes more complicated if the  $A2$  sites are divided into second neighbour and other sites. Note that for both (43.17) and (43.18),  $0 \leq f \leq 1$ , and  $f$  will in general be temperature dependent.

In the early days of diffusion theory a considerable difficulty presented by the vacancy mechanism was that the self-diffusion rate of the solvent in dilute solution is often found to be appreciably greater than that of the pure solvent; in some cases the tracer self-diffusion coefficient may be doubled by the addition of only 1% solute. Since dilute solutions behave ideally so far as the solvent is concerned, we should expect the diffusion coefficient to be given by  $x_{\square}$  multiplied by the diffusion coefficient of vacancies, apart from small correlation

effects, and since  $c_{\square}$  should be nearly the same in dilute solutions as in the pure component, the self diffusion coefficients should be identical. The difficulty is removed by Johnson's proposal, since it is now possible for a faster moving impurity atom to enhance the rate of diffusion of the those solvent atoms which are close to it. This can happen either because  $P$  is greater near to the impurity than it is further away from it (association between the impurity and the vacancy), or because the rate of solvent jumping,  $k_{A1}$  and  $k_{A2}$ , is greater close to the impurity atom than it is in regions remote from it. Analyses of this rather difficult problem have been given by Lidiard (1960) and Howard and Manning (1967).

We finally consider correlation effects in concentrated solutions which have been treated approximately by Manning (1967) with the simplifying assumptions that the rates  $k_A$ ,  $k_B$  at which tracer  $A$  and  $B$  atoms interchange with a vacancy are independent of all the surrounding atoms, and that non-tracer atoms have an average jump frequency of  $x_A k_A + x_B k_B$ . Provided the jump direction is at least an axis of twofold symmetry, he obtains

$$f_B = k_e / (2k_B + k_e), \quad (43.19)$$

where  $k_e$  is the effective frequency for escape of the vacancy from its position adjacent to the tracer  $B$  atom to a random position. Equation (43.19) may be compared with (43.13), (43.17), and (43.18) which contain explicit expressions for  $k_e$  for dilute solutions. In Manning's model, the vacancies are not bound to any particular atoms and

$$k_e = (x_A k_A + x_B k_B) f_{\square} \{2f_0(1-f_0)^{-1}\}, \quad (43.20)$$

where  $f_{\square}$  is the correlation factor for vacancies and  $f_0$  is the correlation factor for self-diffusion in a pure element. Thus the last factor in the equation has values (see p. 397) of 2, 3.77, 5.33, and 7.15 in diamond cubic, simple cubic, b.c.c., and f.c.c. crystals respectively. It remains to give an explicit expression for  $f_{\square}$ , which represents the deviation of the vacancy motion from a random walk. Manning gives

$$f_{\square} = \{x_A k_A (f_A/f_0) + x_B k_B (f_B/f_0)\} / (x_A k_A + x_B k_B) \quad (43.21)$$

so that

$$k_e = 2(x_A k_A f_A + x_B k_B f_B) (1-f_0)^{-1} \quad (43.22)$$

and

$$f_B = \frac{x_A k_A f_A + x_B k_B f_B}{k_B(1-f_0) + x_A k_A f_A + x_B k_B f_B}. \quad (43.23)$$

Note that the vacancies execute a random walk if  $k_A = k_B$ , in which case  $f_A = f_B = f_0$ ;  $f_{\square} = 1$  and (because of the assumptions of the model)  $k_e = 2k_A f_0 (1-f_0)^{-1}$ .

Manning also calculated the effects of the correlations on the intrinsic diffusion coefficients  $D_A$  and  $D_B$ . The non-random vacancy motion gives rise to an additional vacancy flow term, but Darken's relation (41.12) remains valid. However, eqn. (41.16) must be replaced by the following relations between intrinsic and tracer diffusion coefficients:

$$D_A/D_{A+} = (1+x_A B), \quad D_B/D_{B+} = 1-x_B B, \quad (43.24)$$

where

$$B = \frac{(1-f_0)(D_{A+} - D_{B+})}{f_0(x_A D_{A+} + x_B D_{B+})}. \quad (43.25)$$

The vacancy flux may be expressed as

$$-(\mathbf{I}_A + \mathbf{I}_B) = (D_A^* - D_B^*)(\partial g_A / \partial \ln c_A) f_0^{-1} \nabla c_A. \quad (43.26)$$

Since  $f_0^{-1} > 1$ , the Kirkendall shift and the interdiffusion coefficient are larger than predicted by Darken's theory.

As discussed previously, the correlation effects may be represented formally by the off-diagonal elements of the mobility matrix (e.g. eqn. (41.4)). It is therefore interesting to note that with the assumptions made in Manning's theory, these off-diagonal terms obey the Onsager reciprocal relations (41.5).

#### 44. THE THEORY OF DIFFUSION COEFFICIENTS

In this section we consider how the diffusion coefficient is related to the lattice energies. Consider the migration of an  $A$  atom from one position in the structure to a neighbouring vacant position. The initial and final conditions will both be solid solutions containing the same components, and will have the same standard free energies. The free energy will, however, rise to a maximum for some configuration where the atom is not on a regular site, and we may imagine an activated complex to be formed at this stage. The pseudo-thermodynamic theory of Chapter 3 may then be applied to relate the frequency of jumping to the difference in standard free energies of the complex and the equilibrium state. Most theories of diffusion coefficients use this approach, although attempts have been made to develop alternative dynamical theories.

If the solution behaves in an ideal manner, we can write the specific rate constant,  $k_A$  of eqn. (40.4), in the form

$$k_A = (kT/h) \exp(-\Delta g^*/kT), \quad (44.1)$$

where the free energy of activation per atom  $\Delta g^*$  is defined as in eqn. (13.9). From eqn. (41.20) the corresponding tracer diffusion coefficient is

$$D = (Pzr_1^2/6) f(kT/h) \exp(-\Delta g^*/kT). \quad (44.2)$$

Many attempts have been made to apply the reaction rate theory to diffusion in non-ideal solutions. The less satisfactory use eqn. (13.9) and write expressions for  $k_{A,12}$  and  $k_{A,21}$  which include the activity coefficient,  $\gamma_A^*$ , of the activated state. By expanding about the values  $c_A, \gamma_A$  halfway between the compositions of the positions 1 and 2, essentially as in Section 43, an expression for the diffusion coefficient is obtained as

$$D/D_i = (1/\gamma_A^*) \{\gamma_A + c_A(\partial \gamma_A / \partial c_A)\}, \quad (44.3)$$

where  $D_i$  is the value of  $D$  given by eqn. (44.2), and we have made the implicit assumption that  $\gamma_A^*$  is constant. As already emphasized in Chapter 3, it is difficult to give a physical

interpretation to quantities like  $\gamma_A^*$  which appear in equations of this type. Le Claire (1949) has also emphasized that this treatment implies that the mobility of  $A$  depends only on the activity coefficient  $\gamma_A$ . If this is so, the experimental diffusion coefficient defined by Fick's law will vary with composition, but a diffusion coefficient defined in terms of activity gradients will be constant. In a few cases, the available experimental evidence suggests that this condition is nearly satisfied.

As already discussed, these difficulties may be avoided by means of the treatments due to Wert and Zener (1949), and to Vineyard (1957) who introduced only quantities with well-defined physical meanings. For interstitial and vacancy diffusion, the partition function when the diffusing atom is in an equilibrium position is factorized into a function  $Q$ , representing vibration along the path taken in a unit jump process, and a partition function for the remaining degrees of freedom. The tracer diffusion coefficient is then given by

$$D = (Pzr_1^2/6) f(kT/h) (1/Q_\nu) \exp(-\Delta_a g/kT), \quad (44.4)$$

where  $\Delta_a g$  has a clear meaning. It is the isothermal work required to transfer the diffusing atom slowly from its equilibrium position to its saddle-point position, constraining it always to remain in a plane normal to the diffusing path. At the temperatures used in most diffusion experiments,  $Q_\nu$  may be given its classical value  $kT/h\nu$ , where  $\nu$  is some appropriate vibration frequency of the diffusing atom (see p. 88). We thus obtain

$$D = \nu(Pzr_1^2/6) f \exp(-\Delta_a g/kT) = \nu(Pzr_1^2/6) f \exp(\Delta_a s/k) \exp(-\Delta_a h/kT) \quad (44.5)$$

The quantities  $\nu$  and  $f$  which appear in eqn. (44.5) have hitherto been treated as constants, but it is now necessary to examine this assumption. If two different isotopes  $A+$  and  $\alpha+$  are used to measure separate tracer diffusion coefficients,

$$\Delta D/D_{\alpha+} = (\Delta k/k_{\alpha+}) \{1 + (k_{A+}/f_{\alpha+})(\Delta f/\Delta k)\} \simeq (\Delta k/k_{\alpha+}) \{1 + \partial \ln f/\partial \ln k\}, \quad (44.6)$$

where  $\Delta D = D_{A+} - D_{\alpha+}$ , etc. If we use the general form (43.19) for  $f$  and suppose that the non-tracer jumps of frequency  $k_e$  are independent of the isotope, the second bracket in (44.6) becomes simply  $f_{A+}$  and

$$(D_{A+}/D_{\alpha+}) - 1 = f_{A+} \{(k_{A+}/k_{\alpha+}) - 1\}. \quad (44.7)$$

If the activation energy is the same for both isotopes, the ratio of jump probabilities depends only on the attempt frequencies. Then provided the diffusion mechanism involves only one atom which is effectively decoupled from the surrounding atoms, we have (see eqn. (13.23))

$$(k_{A+}/k_{\alpha+}) - 1 = (\nu_{A+}/\nu_{\alpha+}) - 1 = (m_{\alpha+}/m_{A+})^{1/2} - 1. \quad (44.8)$$

This equation is clearly inapplicable to mechanisms such as ring diffusion or interstitialcy diffusion where  $(n-1)$  atoms of the host matrix  $A$  move to new positions at the same time as the jump of the tracer atom. In this case, Vineyard (1957) proposed that eqn. (44.8) should be replaced by

$$(k_{A+}/k_{\alpha+}) - 1 = (\bar{m}_{\alpha+}/\bar{m}_{A+})^{1/2} - 1, \quad (44.9)$$

where

$$n\bar{m}_{A+} = (n-1)m_{A+} + m_{A+}. \quad (44.10)$$

The assumption that the migrating atoms are decoupled from the remainder of the lattice is clearly not valid, and a correction must now be made for the participation of other atoms in the mode which leads to decomposition of the saddle-point configuration. This is expressed (eqn. (13.27)) as

$$(k_{A+}/k_{\alpha+}) - 1 = \Delta K \{ (\bar{m}_{\alpha+}/\bar{m}_{A+})^{1/2} - 1 \}, \quad (44.11)$$

where  $\Delta K$  is the fraction of the translational kinetic energy in the unstable mode (Mullen, 1961; Le Claire, 1966). Equation (44.7) thus becomes in the general case

$$f_{A+} \Delta K = \{ (D_{A+}/D_{\alpha+}) - 1 \} \{ (\bar{m}_{\alpha+}/\bar{m}_{A+})^{1/2} - 1 \}^{-1}. \quad (44.12)$$

The experimental observation that  $D_{A+} \neq D_{\alpha+}$  is known as the isotope effect and its measurement has assumed some importance because of the information it gives on diffusion mechanisms. The ratio of the diffusion coefficients is normally very small but may be measured rather accurately by diffusing both isotopes in the same specimen. Corrections to eqn. (44.12) arise mainly from two of the assumptions made above, namely that there is no mass effect on  $k_e$  and that the activation energies for migration of the two isotopes are equal; the latter is not valid when quantum effects are considered. These corrections have been considered by Le Claire (1966) and Ebisuzaki *et al.* (1967); using the treatment outlined in Section 13, they reached the conclusion that the equation is valid to within a few per cent. In order to extract the maximum information, experimental measurements have thus to be designed to reach this accuracy on the right-hand side.

Although it is not possible in general to separate the product  $f_{A+} \Delta K$  into its separate factors, this may be done in appropriate cases. Thus for tracer self-diffusion in cubic structures, the values of  $f$  are known for the various possible diffusion mechanisms, and the knowledge that  $\Delta K \ll 1$  is then sufficient in some cases to fix the mechanism from the measured values of  $(D_{A+}/D_{\alpha+})$ . For all close-packed metals investigated to date, it is possible in this way to eliminate all but vacancy mechanisms, and the values of  $\Delta K$  range from  $\sim 0.75$  to 1. On the other hand, the results for b.c.c. metals are consistent with all likely mechanisms except four-ring diffusion, and for sodium  $\Delta K = 0.5$  if the monovacancy mechanism is assumed to be operative.

Experimental results for diffusion coefficients are usually written in the form

$$D = D_0 \exp(-\varepsilon/kT), \quad (44.13)$$

and we now see (cf. p. 90) that if  $P$  is constant the experimental activation energy per atom may be identified with  $\Delta_a h$ . For interstitial diffusion,  $P = f = 1$  and the other quantities in eqn. (44.5) have been evaluated (eqns. (40.10) and (40.11)). We thus have

$$\text{and } \left. \begin{aligned} D_0 &= (a^2/6)v \exp(\Delta_a s/k) && \text{(f.c.c. structures),} \\ D_0 &= a^2 v \exp(\Delta_a s/k) && \text{(b.c.c. structures).} \end{aligned} \right\} \quad (44.14)$$

There is a small correction for correlation effects ( $f < 1$ ) if the mechanism is interstitialcy rather than interstitial.

We have already mentioned the proposal by Zener (1951) that  $\Delta_a s$  is necessarily positive. His argument uses the thermodynamic relation

$$\Delta_a s = -(\partial \Delta_a g / \partial T)_p \quad (44.15)$$

to find a correlation between  $\Delta_a s$  and  $\Delta_a h$ . If the free energy of activation is assumed to be due entirely to work done in straining the lattice at the activated complex, one may write  $\Delta_a g / \Delta_a h = \mu / \mu_0$  where  $\mu$  is some appropriate elastic modulus, and  $\mu_0$  is the value of this modulus at absolute zero. Hence

$$\Delta_a s = -\Delta_a h (1/\mu_0) (\partial \mu / \partial T) \quad (44.16)$$

and since  $(\partial \mu / \partial T)$  is always negative,  $\Delta_a s$  should always be positive. Note that if this model is correct, a non-linear dependence of  $\mu$  on  $T$  implies that  $\Delta_a s$ , and hence  $\Delta_a h$ , are temperature dependent quantities.

The possible deficiencies of the elastic model used for calculations of this kind were stated on p. 129. Nevertheless, there is good experimental evidence of a correlation of the type indicated, experimental and theoretical values of  $\Delta_a s$  agreeing well in most cases of interstitial diffusion (Wert, 1950). Zener suggested that where experimental results apparently indicated a negative entropy of activation, the results may be inaccurate because of the unsuspected presence of diffusion short circuits such as grain boundaries, sub-boundaries, or dislocation lines.

Turning now to substitutional diffusion by a monovacancy mechanism, we recall that the quantity  $P$  is equal to the probability  $x_{\square}$  of a given site being vacant. Using eqn. (17.3), we thus find that the experimental activation energy for vacancy self diffusion should be given by

$$\varepsilon = \Delta h_{\square} + \Delta_a h_{\square} = \Delta h_1, \quad (44.17)$$

where  $\Delta_a h$  in eqn. (44.5) has now been equated to the energy of motion of a vacancy  $\Delta_a h_{\square}$ . This relation is at least approximately valid for many f.c.c. metals, but the difficulties in making accurate measurements of  $\Delta h_{\square}$  and  $\Delta_a h_{\square}$  (see Section 17) limit its use as an experimental test for monovacancy diffusion more than was once realized.

Modern techniques have both improved the accuracy and extended the temperature range of diffusion measurements, and have thus permitted more detailed interpretation of  $\ln D$  vs.  $1/T$  plots. Although such plots are frequently almost linear for f.c.c. metals, some curvature can nevertheless be detected and there are strong indications that this is due, at least partially, to a divacancy contribution to self-diffusion. Thus it may be misleading to use a least squares linear fit to the experimental data in conjunction with eqn. (44.17) and we now consider how the measured values of  $D$  may be divided into monovacancy and divacancy contributions.

If diffusion involves monovacancies and divacancies, the self-diffusion coefficient may be written

$$D = D_{01} \exp(-\Delta h_1/kT) + D_{02} \exp(-\Delta h_2/kT), \quad (44.18)$$

where  $\Delta h_2$  for divacancies is given by an expression similar to (44.17) and the pre-exponential factors (see eqn. (44.5)) are for f.c.c. structures.

$$D_{01} = a^2 v_1 f_1 \exp\{(\Delta s_{\square} + \Delta_a s_{\square})/k\} = a^2 v_1 f_1 \exp(\Delta s_1/k) \quad (44.19)$$

and

$$D_{02} = 4a^2 v_2 f_2 \exp(\Delta s_2/k). \quad (44.20)$$

The numerical factors in this equation arise as follows. The Einstein-Smoluchowski factor eqn. (40.7) is equal to  $a^2$  for monovacancy diffusion in f.c.c. (and b.c.c.) structures and is equal to  $4a^2/6$  for divacancy diffusion in f.c.c. on the assumption that the moving atom is always a nearest neighbour of both vacancies; any other jump will decompose the divacancy. The probability  $P$  for divacancies contains a factor  $z/2 = 6$  representing the number of orientations of the divacancy in the crystal (see p. 130). Note that the Einstein-Smoluchowski factor for diffusion of the divacancies themselves is  $\frac{1}{6}$ .

Eqn. (44.18) contains four disposable parameters and may thus be fitted to curved  $\ln D$  vs.  $1/T$  plots as was done for nickel by Seeger *et al.* (1965) and for gold by Wang *et al.* (1968). However, alternative interpretations may be based on temperature dependencies of the activation enthalpies and entropies, and it seems necessary to allow for this possibility. Seeger and Schumacher (1967) considered that it is sufficient to assume that the activation energy varies linearly with temperature so that

$$\Delta h_1 = \Delta h_{10} + C_7(T - T_0), \quad (44.21)$$

where  $\Delta h_{10}$  is the value at a reference temperature  $T_0$  and  $C_7$  is a constant which combines the separate temperature dependences of  $\Delta h_{\square}$  and  $\Delta_a h_{\square}$ . Provided the temperature  $T_0$  is above the Debye temperature  $\theta_D$ , the second term in (44.14) is small in comparison with the first. The corresponding temperature dependence of the entropy is now

$$\Delta s_1 = \Delta s_{10} + C_7 \ln(T/T_0) \quad (44.22)$$

and eqn. (44.18) may be written in the form

$$\begin{aligned} \ln D = \ln D_{01} - (\Delta h_{10}/kT) + \ln\{1 + D_{02}/D_{01}\} \exp(\Delta h_{10} - \Delta h_{20})/kT \\ + C_7 \ln(T/T_0) - C_7(T - T_0)/kT, \end{aligned} \quad (44.23)$$

where it has been assumed that the temperature dependences of  $\Delta h_2$ ,  $\Delta s_2$  are equivalent to those of  $\Delta h_1$ ,  $\Delta s_1$ . The temperature variation of  $\Delta h_2$  is a second-order correction and this assumption is unimportant.

Equation (44.23) contains five independent parameters, and since the observed deviation from linearity is small, an analysis of experimental results in terms of this equation alone cannot readily be made. However, additional information may also be employed; for example, experimental determinations of the correlation factor from the isotope effect or from a comparison of tracer diffusion with macroscopic thermal or electron diffusion will enable  $D_{02}/D_{01}$  to be deduced. This procedure is based on the knowledge that the correlation factor for single vacancies is  $\sim 1.67$  times larger than that for divacancies. The measured correlation factor is an effective factor which varies with the relative contributions

of the two mechanisms, and hence with temperature. Mehrer and Seeger (1969) used the results of Rothman and Peterson (1969) on the isotope effect in copper to analyse the self-diffusion data which were originally represented by

$$D = 7.8 \times 10^{-5} \exp(-2.19 \text{ eV}/kT) \text{ m}^2 \text{ s}^{-1} \quad (44.24)$$

into a divacancy contribution and a single-vacancy contribution which is

$$D_1 = 1.9 \times 10^{-5} \exp(-2.09 \text{ eV}/kT) \text{ m}^2 \text{ s}^{-1} \quad (44.25)$$

at room temperature. When combined with results for  $\Delta h_{\square}$  obtained by Simmons and Balluffi (1963) (see Section 17), this gives

$$\Delta h_{\square} = 1.04 \text{ eV}, \quad \Delta_a h_{\square} = 1.05 \text{ eV},$$

whilst eqn. (44.17) is not valid if  $\varepsilon$  is taken from (44.24).

A related procedure for analysing eqn. (44.23) is to make use of data on the pressure dependence of the diffusion rate which determines an effective activation volume at each temperature. Since the activation volumes for vacancies and divacancies are different, this again enables their relative contributions to  $D$  to be assessed (e.g. Mehrer and Seeger, 1972). The expressions for the effective activation volume are rather complex, but this method is possibly more accurate than the isotope effect procedure since the activation volume of a divacancy is approximately twice that of a monovacancy. Finally, nuclear magnetic resonance techniques enable diffusion coefficients to be measured at temperatures so low that the divacancy contribution is negligible; combination of these results with high temperature measurements then gives both the vacancy and divacancy parts of  $D$ .

We now turn to consider some of the special difficulties which are encountered in explaining diffusion measurements on b.c.c. metals. For most metals, self-diffusion appears to have the following characteristics (Le Claire, 1965); (1) the Arrhenius law (eqn. (44.13)) is approximately valid, (2) the activation energy  $\varepsilon$  is given within  $\sim 20\%$  by the empirical correlations

$$\varepsilon \approx 34T^{sl} \approx 16.5 \Delta h^{sl} \quad (44.26)$$

where  $T^{sl}$  and  $\Delta h^{sl}$  are respectively the melting temperature and the latent heat of melting, and (3) the pre-exponential factor  $D_0$  is between  $5 \cdot 10^{-6}$  and  $5 \cdot 10^{-4} \text{ m}^2 \text{ s}^{-1}$ . Some b.c.c. metals conform to these general rules, albeit with rather larger curvatures in their Arrhenius plots than are usually found for f.c.c. metals, but other metals have markedly non-linear Arrhenius plots.

Analyses of the divacancy contribution to diffusion in b.c.c. metals are more complex than in f.c.c. metals because motion of a nearest neighbour divacancy by means of nearest neighbour atom jumps necessarily implies the temporary dissociation of the divacancy. There are, of course, other possibilities; for example, the stable divacancy configuration may be two next nearest neighbour vacant sites (p. 138), or atomic jumps may take place between next nearest neighbour sites. The possibility of associative and dissociative jumps means that the correlation factor  $f$  for divacancies is temperature-dependent in the b.c.c. structure, and Seeger (1972) has suggested that the most stable configurations and corres-

diffusion coefficient. Lattice diffusion is important only because the majority of the atoms in a crystal are unable to utilize the more rapid diffusion paths.

In principle, the measured diffusion coefficient varies with grain size but the contribution of the grain boundaries to the overall diffusion rate is very small under the conditions of most diffusion anneals. Analyses of grain boundary diffusion are thus usually made by studying experimental concentration contours, which show greatest penetration along the boundaries. At very low temperatures where  $(Dt)^{1/2}$  is less than an interatomic distance  $r_1$ , diffusion will be strictly confined to grain boundaries and dislocations, and penetration will still be mainly concentrated around the boundaries if  $r_1 < (Dt)^{1/2} < L^B$ , the grain diameter. However, if  $(Dt)^{1/2} > L^B$ , a diffusing atom will encounter several grain boundaries in time  $t$ , so that the boundary fields overlap and the overall diffusion can be described by a weighted average diffusion coefficient (see below). These three regimes were called types *C*, *B* and *A* respectively by Harrison (1961) and Gupta *et al.* (1978).

A difference in diffusion behaviour is to be expected for low-angle and high-angle grain boundaries. In a high-angle boundary, diffusion should be isotropic with a much reduced activation energy, whereas in low-angle grain boundaries, the geometry of the dislocation arrays will influence the diffusion. In a symmetrical low-angle tilt boundary, for example, diffusion along the edge dislocations will be rapid, and a diffusion coefficient for each individual dislocation pipeline may be envisaged. For diffusion normal to the dislocation lines, the atoms must cross coherent regions of boundary, and the activation energy will not be very different from that for lattice diffusion. The boundary diffusion coefficient will thus be anisotropic in such a boundary.

Experimental measurements of grain boundary diffusion coefficients from observed concentration plots depend upon an analysis first given by Fisher (1951), who used the model shown in Fig. 9.3. The grain boundary is assumed to be a uniform slab of thickness  $\delta^B$  having a diffusion coefficient  $D^B$  which is much greater than the lattice diffusion coefficient,

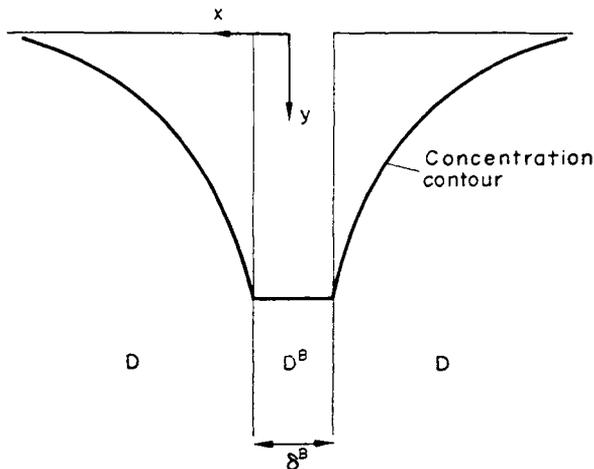


FIG. 9.3. To illustrate Fisher's model for grain boundary diffusion.

ponding jump mechanisms may change from metal to metal. This could explain the appreciably greater variability in the experimental diffusion results for "normal" b.c.c. metals when compared with those for f.c.c. metals.

The "anomalous" b.c.c. metals include titanium, zirconium, hafnium and probably praeosdymium, uranium, cerium, and plutonium, all of which have phase transitions on cooling from a b.c.c. structure to some other structure. When measurements are available over sufficiently large ranges of temperature, the resultant Arrhenius plots are very curved, and if these are approximated as two straight lines the lower temperature part of the curve indicates values of  $\epsilon$  and  $D_0$  much lower than are indicated by the above rules for normal metals. Although various explanations have been proffered in terms of diffusion short circuits or strong impurity-vacancy binding, the parameters required to fit the observations are implausible, and there also seems no reason why particular metals should be different in these respects from the other b.c.c. metals. An alternative suggestion (Seeger, 1972) is that the anomalous behaviour is linked to the phase transition, near to which there is an abnormally large temperature variation of the monovacancy parameters  $\Delta h_1$  and  $\Delta s_1$ . The curvature of the Arrhenius plot is thus attributed mainly to this temperature dependence, although at high temperatures there will also be a divacancy contribution.

Vanadium and chromium were originally included in the anomalous metals, but although they exhibit rather marked curvatures in their  $\ln D$  vs.  $1/T$  plots, the deviation from normal behaviour is now at high temperatures, where  $D_0$  and  $\epsilon$  are apparently higher than usual. Seeger suggests that these metals differ only quantitatively from the normal b.c.c. metals in that they have a rather larger divacancy contribution which is characterized by an unusually large entropy factor.

The discussion in this section is mainly concerned with self-diffusion coefficients. Interstitial impurities usually diffuse at a much faster rate than the solvent atoms, but the impurity diffusion coefficients for substitutional impurities are often of the same magnitude as those of the host atoms. However, some substitutional impurities also diffuse at a very fast rate; these include noble metal solutes in silicon, germanium, lead, tin, thallium, etc. (Anthony, 1970). The explanation of this effect was first given by Frank and Turnbull (1956); the solute atoms dissolve both substitutionally and interstitially, and although the interstitial solubility may only be of the order of one-hundredth of the substitutional solubility, the interstitial atoms nevertheless make the biggest contribution to the diffusion rate. In effect, the diffusion mechanism involves the dissociation of a substitutional solute atom into an interstitial atom plus a vacancy.

#### 45. STRUCTURE SENSITIVE DIFFUSION PROCESSES

In Sections 39-44 we have discussed diffusion through regions of crystal which are reasonably perfect except for small concentrations of defects. We expect that the activation energy for diffusion will be much lower in regions of bad crystal, so that diffusion should be a "structure sensitive" phenomenon. Under appropriate circumstances, we may define diffusion coefficients for the transfer of atoms along free surfaces, grain boundaries, or individual dislocation lines, and we expect that all of these will be larger than the lattice

*D.* The concentration of diffusing substance at the original interface  $c_0$  is assumed to remain constant with time, and the equilibrium concentration in the boundary is assumed to be the same as that in the lattice. The concentration in the boundary is also supposed to be uniform across its width, and both diffusion coefficients are regarded as independent of concentration. Finally, Fisher made the simplifying assumption that direct diffusion from the original interface into the lattice can be disregarded in comparison with diffusion along the boundary and then out sideways into the lattice. With these rather restrictive conditions, an approximate solution for the concentration at any point  $x, y$  (Fig. 9.3) at time  $t$  can be written in the form

$$c = c_0 \exp\left\{-\frac{2D}{D^B \delta^B}^{1/2} / (\pi Dt)^{1/4}\right\} y \operatorname{erfc}\left\{x/2(Dt)^{1/2}\right\}. \quad (45.1)$$

By integrating this equation, the following relation is obtained between the mean concentration of diffusing material in a thin slab at a distance  $y$  from the original interface and the distance  $y$

$$\bar{c} \propto t^{1/2} \exp\left\{-\frac{2D}{D^B \delta^B}^{1/2} / (\pi Dt)^{1/4}\right\} y + \text{const.} \quad (45.2)$$

so that a plot of  $\ln \bar{c}$  against  $y$  should give a straight line, from the slope of which  $\delta^B D^B/D$  may be determined. Note that instead of the Gaussian penetration profile characteristic of volume diffusion (p. 388), the concentration falls off exponentially as the first power of the penetration distance  $y$ .

Fisher's assumption that the concentration change in the crystal due to direct diffusion through the lattice from the original interface may be neglected simplifies the problem, since it implies that lattice diffusion is everywhere normal to the grain boundary. It is clear, however, that this assumption is not really justified, and a more complete treatment of the Fisher model by Whipple (1954) manages to avoid it. The initial and boundary conditions of the model are  $c(x, y, 0) = 0$  and  $c(0, 0, t) = c_0$ , and the diffusion equation has to be solved inside and outside the grain boundary slab in such a way that the concentration  $c$  and the diffusion flux  $D(\partial c/\partial x)$  are both continuous at the surface of the slab. Whipple obtains a result for the concentration at any point  $c(x, y)$  as a rather complex integral function which we shall not reproduce here. The form of the result is different from Fisher's inasmuch as there is no longer a linear relation between the mean concentration and the penetration distance, but Turnbull and Hoffman (1954) show that in a typical case two treatments give nearly identical results.

Le Claire (1951) suggested that a simpler way of obtaining the grain boundary diffusion coefficient is to measure the angle  $\psi$  at which a curve of constant composition meets the grain boundary. This angle is equal to  $(dx/dy)_{x=0}$ , and a simple differentiation of eqn. (45.1) gives

$$\delta^B D^B/D = 2(\pi Dt)^{1/2} \cot^2 \psi. \quad (45.3)$$

Although Whipple's solution does not give a simple analytical expression like this, he has computed an analogous relation. An important result emphasized by Whipple is that the depth of penetration of a particular concentration contour along a grain boundary is not

markedly greater than the depth of penetration into the grain, even with large ratios of  $D^B/D$ . This is because most of the material which originally diffuses down the boundary is lost sideways into the lattice. Typical figures are that a ratio of  $D^B/D$  of about  $10^5$  is needed to produce a penetration depth along the boundary twice as large as that in the crystal in about 100 h. It will be realized from the above discussion that all grain boundary diffusion experiments depend upon  $D^B\delta^B$ ; when it is desired to write values for  $D^B$  it is usual to assume a thickness of two or three atom diameters for the boundary region.

This analysis of grain boundary diffusion would be expected to apply most successfully to high angle boundaries, where the structure is very disordered, but Turnbull and Hoffman (1954) showed that it is also valid for the self-diffusion of silver along low-angle (dislocation) boundaries. Using symmetrical tilt boundaries, these workers found that the overall diffusivity varied with the misorientation  $\theta$  as  $\sin \frac{1}{2}\theta$ . Since the density of dislocations in such a boundary is also proportional to  $\sin \frac{1}{2}\theta$  (p. 336), the results imply that it is possible to define an intrinsic diffusion coefficient  $D^F$  along each dislocation pipe, the diffusion down one dislocation being unaffected by the other dislocations. The relation between the boundary diffusion coefficient and the dislocation diffusion coefficient is then

$$D^B = 2D^F a \sin \frac{1}{2}\theta \quad (45.4)$$

for the simple boundary considered.

The activation energy for self-diffusion along dislocation pipes obtained in these and similar experiments, and the activation energy for diffusion in high-angle grain boundaries, are both found to be of the order of half the corresponding activation energy for lattice diffusion. This is not unexpected if the core of a dislocation and a high-angle boundary have similar structures, and several other experiments indicate that the diffusivity of a dislocation and a general grain boundary are similar. Results on chemical diffusion along grain boundaries are scarcer, but the activation energy is again probably of the order of half that for lattice diffusion. Chemical diffusion will be complicated, amongst other effects, by a tendency for the solute atoms to segregate preferentially either in or away from the boundaries.

In macroscopic diffusion experiments, the grain boundaries will make increasing contributions at low temperatures. The boundary mechanism will be dominant when the ratio of the diffusion coefficients exceeds the ratio of the cross-sectional areas of the respective paths, which is about  $10^5$  for a grain size of 0.1 mm. This gives a temperature of about one-half to three-quarters of the melting point temperature for the metals for which data are available. The transition will be gradual, and may account for some apparently incorrect early results on the temperature variation of diffusion coefficients. For greater accuracy in the determination of experimental activation enthalpies, it is normally recommended that the temperature range covered should be as large as possible. This is only true, of course, if the mechanism remains unchanged; unfortunately the scatter of diffusion data is such that apparently good Arrhenius type plots may be made even when the diffusion process changes with falling temperature from mainly lattice to mainly grain boundary. Many authors have emphasized the danger that incorrect activation energies and enthalpies may result from this kind of effect in polycrystalline specimens (see p. 410).

The enhanced diffusion rate along individual dislocation lines suggests the possibility that a grown-in three-dimensional dislocation network may make an appreciable contribution to the diffusion rate in a single crystal. This situation has been considered by Hart (1957), using a random walk type of model for the diffusion process. The mean square displacement of an atom in a given time may be regarded as partly due to migration through the lattice and partly to much more rapid migration along dislocation lines. Provided the dislocation network has connecting lines of length short compared to the diffusion penetration distance, the penetration profile will be Gaussian as for true random volume diffusion, but the apparent diffusion coefficient is larger than the true volume diffusion coefficient.

Suppose an atom moves randomly through the crystal until it encounters a dislocation, and that it then remains for a time  $\tau^D$  at the dislocation core, during which time it diffuses a mean square distance  $\overline{P^2}$ . If this procedure is repeated several times during a diffusion period  $t$ , the total mean square migration distance (averaged for many atoms) may be taken as the sum of the mean square migration distance ( $\overline{Q^2}$ ) through the lattice in the time  $t$ , and the mean square migration distances along dislocation lines, giving

$$\overline{S^2} = m\overline{P^2} + \overline{Q^2}, \tag{45.5}$$

where  $m$  is the number of dislocations encountered during the time  $t$ . Correlation effects, which are ignored, should be small.

If the temperature and the time  $t$  are such that the mean migration time of an atom between two dislocations is much less than  $t$ , it is justifiable to assume that a diffusing atom spends a fraction  $x^D$  of the time diffusing along dislocations and a fraction  $(1 - x^D)$  of the time in the lattice, where  $x^D$  is the fractional concentration of atoms in dislocation core regions. This gives

$$m\tau^D = x^D t, \tag{45.6}$$

and the lattice diffusion coefficient  $D$  is given by

$$\overline{Q^2} = 6D(1 - x^D)t \tag{45.7}$$

(see eqn. (40.5)).

From eqns. (45.5), (45.6), and (45.7) we now have the effective diffusion coefficient  $D'$  defined by<sup>†</sup>

$$\overline{S^2} = 6D't,$$

where

$$D' = (\overline{P^2}/6\tau^D)x^D + D(1 - x^D). \tag{45.8}$$

The term  $(\overline{P^2}/6\tau^D)$  is probably not much different from the diffusion coefficient for dislocation pipes in low angle boundaries. We have seen on p. 314 that a reasonable estimate of  $x^D$  is probably about  $10^{-7}$  for well-annealed material. Thus if the grain boundary diffusion

<sup>†</sup> In Hart's paper, the factor 6 in eqns. (45.7) and (45.8) is replaced by a factor 2; this is because he considers the mean square displacements projected along a particular direction. This gives  $\overline{S_i^2} = 2Dt$  for the diffusion coefficient in that direction, but reduces to (40.5) if the diffusion coefficient is isotropic (see p. 386).

coefficient is  $10^6$  times larger than the bulk lattice diffusion coefficient, the dislocation network could make a 10% contribution to the overall diffusion flux. A ratio of this magnitude is quite possible in low-temperature diffusion experiments, and may explain deviations from the Arrhenius plot (i.e. discrepancies between high and low-temperature diffusion results) even in single crystals. At sufficiently low temperatures, the dislocations may make much more significant contributions to diffusion than the lattice, but the overall diffusion rate is then so small that measurements are extremely difficult. Hart's derivation of a mean diffusion coefficient is clearly also applicable to combined lattice and grain boundary diffusion in Harrison's type *A* situation, and the ratio of the number of atoms in the boundaries to the number in the lattice is then  $\sim \delta^B/L^B$ , so that the effective diffusion coefficient becomes

$$D' = D + \delta^B D^B / L^B. \quad (45.8a)$$

Cahn and Balluffi (1979) have considered the modifications produced if the grain boundaries are migrating with velocity  $Y$  during the diffusion. The type *A* (or multiple boundary diffusion) regime then requires  $(Dt)^{1/2} + Yt > L^B$  and eqn. (45.8a) is valid even if  $D$  is negligible. When this condition is not satisfied, penetration is mainly isolated around grain boundaries, and Cahn and Balluffi distinguish four further cases, two of which (*C* and *B*) have stationary boundaries ( $Yt < r_1$ ) and respectively no lattice diffusion and some lattice diffusion (transition at  $Dt = r_1^2$ ) whilst the other two having moving boundaries ( $Yt > r_1$ ) but are sub-divided according to whether or not there is lattice diffusion ahead of the moving boundary (transition along the line  $Yr_1 = D$ ). The classical Fisher-Whipple analysis applies to region *B* and requires modification if the boundary is migrating. All the transitions described above are gradual, and the conditions given are only approximate.

We now turn to consider the effects of non-equilibrium states on the rate of diffusion. We might expect enhanced diffusion in a severely deformed metal because of an increase in the number of dislocations or of vacancies, or in quenched or irradiated materials, because of increases in the vacancy concentration over the thermal equilibrium value. In normal diffusion anneals at reasonably high temperatures, the excess defect concentrations will be removed during the anneal, so that the amount of extra diffusion obtained will depend on the lifetime of the defects considered. More effective increases in diffusion rates should be obtained if defects are continuously created whilst the diffusion is taking place, e.g. by irradiating the specimen, or maintaining a stress on it. We now consider these effects more quantitatively.

For vacancy diffusion the diffusion coefficient is directly proportional to the atomic concentration  $x_{\square}$  of vacancies. As we have seen in Section 17, this can be as high as  $10^{-3}$  to  $10^{-4}$  in severely quenched or damaged metals, so that at low temperatures the increase in diffusion coefficient can be a very large factor. The discussion on vacancy annealing in Section 17, however, shows that this is not an easy effect to detect in normal experiments. Thus Lomer (1958) points out that if a vacancy makes  $n$  jumps on the average before disappearing, the average number of jumps made by an atom is  $x_{\square}n$  and this is equal to  $6Dt/r_1^2$  (eqns. (40.5) and (40.7)). The distance  $2(Dt)^{1/2}$  which defines the mean diffusion length is thus

given by

$$2(Dt)^{1/2} = r_1(2nx_{\square}/3)^{1/2}, \quad (45.9)$$

and with typical values this is about one micron if  $n$  is taken as  $10^{10}$  (see p. 140). Thus although the excess vacancies would be extremely effective in promoting diffusion over short distances, they anneal out too quickly to be important in macroscopic diffusion measurements. Effects of this kind, however, are probably extremely important in such transformation phenomena as the ageing of supersaturated solid solutions in which initial segregation over very short distances almost certainly utilizes the excess vacancies retained in the quenched specimen.

Similar arguments have been given by Lomer to show that the effect of extra vacancies produced by continuous irradiation is also rather small, especially since the effective number of jumps made by a vacancy before disappearing is usually much less in irradiated specimens than in quenched specimens (see Section 17). The general conclusion is again that only at low temperatures do the contributions from the excess vacancies begin to outweigh those from the thermal vacancies which are always present, and that the excess vacancies are only important for fine scale phenomena.

Finally, we discuss the effect of plastic deformation on diffusion rate. A number of workers have reported very considerably enhanced diffusion rates in metals subjected to torsional or compressive strains during the diffusion process, the increase being of the order of a hundred or a thousand fold in some circumstances. Other workers have reported no change in diffusion rates, even for large total strains and high strain rates. This subject is thus rather confused and controversial and, in particular, there is disagreement about the magnitude of the effect in silver at about  $800^{\circ}\text{C}$ . It seems possible that genuine large increases are obtained in some circumstances, especially at lower temperatures, but the mechanism remains uncertain. Some workers have attributed the effect to point defects generated by moving dislocations, but the arguments above seem to suggest that very long vacancy lifetimes would be required for this model, and there is no available evidence to support such long lifetimes at high temperatures. However, the suggestion that "old" dislocations cannot absorb vacancies readily may be relevant here, since it suggests that sinks for vacancies will be less effective at diffusion temperatures. Annihilation of vacancies by interstitials will not be so important in deformation experiments as in irradiation, since present evidence is that moving dislocations generate vacancies but not interstitials. The alternative possibility is that enhanced diffusion is due to greatly increased sub-boundary diffusion, especially if the sub-boundaries are themselves migrating during the experiment.

## REFERENCES

- ANTHONY, T. R. (1970) *Vacancies and Interstitials in Metals*, p. 935, North-Holland, Amsterdam.  
 BAKKER, H. (1971) *Phys. stat. sol.* **44**, 369.  
 BALLUFFI, R. W. (1954) *Acta metall.* **2**, 194.  
 BARDEEN, J. and HERRING, C. (1951) *Atom Movements*, p. 87., American Society for Metals, Cleveland, Ohio; (1952) *Imperfections in Nearly Perfect Crystals*, p. 261, Wiley, New York.  
 BARNES, R. S. (1952) *Proc. Phys. Soc. B*, **65**, 512; (1955) *Defects in Crystalline Solids*, p. 359, The Physical Society, London.

- BARNES, R. S. and MAZEY, D. J. (1958) *Acta metall.* **6**, 1.  
 BECKER, R. (1937) *Z. Metallk.* **29**, 245.  
 CAHN, J. W. (1961) *Acta metall.* **9**, 795; (1962) *Ibid.* **10**, 179.  
 CAHN, J. W. and BALLUFFI, R. W. (1979) *Scripta metall.* **13**, 499.  
 CARSLAW, H. S. and JAEGER, J. C. (1947) *Conduction of Heat in Solids*, Clarendon Press, Oxford.  
 COLEMAN, B. D. and TRUESDELL, C. (1960) *J. Chem. Phys.* **33**, 28.  
 COMPAAN, K. and HAVEN, Y. (1958) *Trans. Faraday Soc.* **54**, 1498.  
 COOK, H. E. and HILLIARD, J. E. (1969) *J. Appl. Phys.* **40**, 2191.  
 CRANK, J. (1956) *The Mathematics of Diffusion*, Clarendon Press, Oxford.  
 DARKEN, L. S. (1948) *Trans. Am. Inst. Min. (Metall.) Engrs.* **175**, 184.  
 DYSON, B. F., ANTHONY, J., and TURNBULL, D. (1966) *J. Appl. Phys.* **37**, 2370; (1967) *Ibid.* **38**, 3408; (1968) *Ibid.* **39**, 1391.  
 EBISUZAKI, Y., KASS, W. J., and O'KEEFFE, M. (1967) *J. Chem. Phys.* **46**, 1373; *Phil. Mag.* **15**, 1071.  
 EINSTEIN, A. (1905) *Ann. Physik.* **14**, 549.  
 FISHER, J. C. (1951) *J. Appl. Phys.* **22**, 74.  
 FRANK, F. C. and TURNBULL, D. (1956) *Phys. Rev.* **104**, 617.  
 GUPTA, D., CAMPBELL, D. R. and HO, P. S. (1978) in *Interdiffusion and Reactions*, p. 161, Wiley, New York.  
 HARRISON, L. G. (1961) *Trans. Faraday Soc.* **57**, 1191.  
 HART, E. W. (1957) *Acta metall.* **5**, 597.  
 HEUMANN, T. and WALTHER, G. (1957) *Z. Metallk.* **48**, 151.  
 HILLERT, M. (1961) *Acta metall.* **9**, 525.  
 HOWARD, R. E. (1966) *Phys. Rev.* **144**, 650.  
 HOWARD, R. E. and LIDIARD, A. B. (1964) *Reports Prog. Phys.* **27**, 161.  
 HOWARD, R. E. and MANNING, J. R. (1967) *Phys. Rev.* **154**, 561.  
 HUNTINGTON, H. B. (1951) *Atom Movements*, p. 69, American Society for Metals, Cleveland, Ohio.  
 JOHNSON, W. A. (1941) *Trans. Am. Inst. Min. (Metall.) Engrs.* **143**, 107.  
 KIRKALDY, J. S. (1985) *Scripta Met.* **19**, 1307  
 KIRKALDY, J. S. and YOUNG, D. J. (1987) *Diffusion in the Condensed State*, the Institute of Metals, London.  
 KONOBEVSKY, S. T. (1932) *J. Exp. Theor. Phys. USSR* **13**, 185; (1943) *Ibid.* **13**, 418.  
 LAZARUS, D. (1960) *Solid State Phys.* **10**, 71.  
 LE CLAIRE, A. D. (1949) *Prog. Metal Phys.* **1**, 306; (1951) *Phil. Mag.* **42**, 468; (1953a) *Prog. Metal Phys.* **4**, 265; (1953b) *Acta metall.* **1**, 438; (1965) *Diffusion in Body Centred Cubic Metals*, p. 3, American Society for Metals, Metals Park, Ohio; (1966) *Phil. Mag.* **14**, 1271.  
 LE CLAIRE, A. D. and LIDIARD, A. B. (1956) *Phil. Mag.* **1**, 518.  
 LIDIARD, A. B. (1960) *Phil. Mag.* **5**, 1171.  
 LOMER, W. M. (1958) *Vacancies and Other Point Defects in Metals and Alloys*, p. 79, The Institute of Metals, London; (1959) *Prog. Metal Phys.* **8**, 255.  
 MANNING, J. R. (1962) *Phys. Rev.* **128**, 2169; (1964) *Ibid.* **136**, A1758; (1967) *Acta metall.* **15**, 817; (1968) *Diffusion Kinetics for Atoms in Crystals*, Van Nostrand, Princeton, New Jersey.  
 MATANO, C. (1933) *Japan. J. Phys.* **8**, 109.  
 MEHRER, H. (1972) *J. Phys. (F)* **2**, L11.  
 MEHRER, H. and SEEGER, A. (1969) *Phys. stat. sol.* **35**, 313; (1972) *Crystal Lattice Defects* **3**, 1.  
 MULLEN, J. G. (1961) *Phys. Rev.* **121**, 1649.  
 NACHTRIEB, N. H. and HANDLER, G. S. (1954) *Acta Metall.* **2**, 797.  
 ONSAGER, L. and FUOSS, R. M. (1932) *J. Phys. Chem.* **36**, 2689.  
 PANETH, H. (1950) *Phys. Rev.* **80**, 708.  
 PETERSON, N. L. (1968) *Solid State Phys.* **22**, 409.  
 PHILOFSKY, E. M. and HILLIARD, J. E. (1969) *J. Appl. Phys.* **40**, 2198.  
 RESNICK, R. and SEIGLE, L. L. (1957) *Trans. Am. Inst. Min. (Metall.) Engr.* **209**, 87.  
 RICE, S. A. (1958) *Phys. Rev.* **112**, 804.  
 ROTHMAN, S. J. and PETERSON, N. L. (1969) *Phys. stat. sol.* **35**, 305.  
 SEEGER, A. (1972) *J. Less-common Metals* **28**, 387.  
 SEEGER, A. and SCHUMACHER, D. (1967) *Mater. Sci. Engr.* **2**, 31.  
 SEEGER, A., SCHOTTKY, G., and SCHUMACHER, D. (1965) *Phys. stat. sol.* **11**, 363.  
 SEITZ, F. (1948) *Phys. Rev.* **74**, 1513; (1950a) *Acta crystallogr.* **3**, 355; (1950b) *Phase Transformations in Solids*, p. 77, Wiley, New York; (1953) *Acta metall.* **1**, 355.  
 SHEWMON, P. G. (1963) *Diffusion in Solids*, McGraw-Hill, New York.

- SHIRN, G. A., WAJDA, E. S., and HUNTINGTON, H. B. (1953) *Acta metall.* **1**, 513.  
SIMMONS, R. O. and BALLUFFI, R. W. (1963) *Phys. Rev.* **129**, 1533.  
SMIGELSKAS, A. D. and KIRKENDALL, E. O. (1947) *Trans. Am. Inst. Min. (Metall.) Engrs.* **171**, 130.  
TURNBULL, D. and HOFFMAN, R. E. (1954) *Acta metall.* **2**, 419.  
VINEYARD, G. H. (1957) *J. Phys. Chem. Solids* **3**, 121.  
WANG, C. G., SEIDMAN, D. N., and SIEGEL, R. W. (1968) *Phys. Rev.* **169**, 553.  
WERT, C. (1950) *Phys. Rev.* **79**, 601.  
WERT, C. and ZENER, C. (1949) *Phys. Rev.* **76**, 1169.  
WHIPPLE, R. T. P. (1954) *Phil. Mag.* **45**, 1225.  
ZENER, C. (1950) *Acta crystallogr.* **3**, 346; (1951) *J. Appl. Phys.* **22**, 372; (1952) *Imperfections in Nearly Perfect Crystals*, p. 305, Wiley, New York.

## CHAPTER 10

# *The Classical Theory of Nucleation*

### 46. THE FORMATION OF NUCLEI OF A NEW PHASE

The kinetics of a heterogeneous reaction, as discussed in Section 4, can usually be described in terms of the separate nucleation and growth of the transformed regions. The classical theory of nucleation by random fluctuations in a metastable assembly is due mainly to Volmer and to Becker and Döring, but many other workers have made significant contributions. This theory was formulated first for the simplest nucleation process, the condensation of a pure vapour to form a liquid, and we shall find it convenient to study this change in some detail as a prelude to the more complex problems of nucleation in liquid and solid phases. Throughout this chapter we deal only with transformations which do not involve changes of composition; the complications which these introduce are considered in later chapters.

The essential driving force for a phase transformation is the difference in the free energies of the initial and final configurations of the assembly, but when small particles of the new phase are formed, the free energy rises at first. The increase is due to the considerable proportion of atoms in these particles which are situated in transition regions between the phases, where they do not have the environment characteristic of the new phase in bulk. The situation is conventionally described by assigning volume free energies to the bulk phases, and a surface free energy to the interface region; for sufficiently small particles of the new phase, the surface term is dominant. Additional factors have also to be considered in transformations in the solid state, where the volume change associated with the transformation, and the possible tendency for the two phases to remain coherent across the interphase boundary, may both produce considerable perturbations of the atomic arrangement. Each nucleus may then be regarded as a source of internal stress, and the resultant elastic strain energy has to be included in the overall free energy change of the assembly.

We begin with the vapour-liquid transformation where there are no strain energies to be considered. The usual theory is developed in terms of a model of a very small liquid droplet, which is certainly invalid; the justification for its use is partly mathematical, since we shall find that the assumptions are least important where they are most outrageous and partly empirical, since reasonable agreement with experiment is obtained. The essential assumption is that a small droplet may be treated in the same way as a large mass of liquid, and in particular may have its properties described by macroscopic thermodynamic parameters.

Thus we treat the transition region between vapour and liquid as a geometrical surface with a specific free energy  $\sigma$  per unit area. As discussed in Section 35, the value of  $\sigma$  for a curved boundary will vary with the position chosen for the geometrical surface, but no serious ambiguity arises unless the radius of curvature is of the order of the width of the transition region. Unfortunately, this is just the situation encountered in nucleation theory, where the rate of nucleation is commonly determined by the properties of droplets about twenty atomic diameters or less across. The concept of surface free energy is thus scarcely appropriate to the problem, and the value of  $\sigma$  for any particular choice of reference interface will not necessarily bear any simple relation to the usual macroscopic free energy  $\sigma_\infty$  defined for a plane interface. Similar difficulties arise in connection with the other parameters used to define the droplet. In the transition zone, estimated at two to seven atomic or molecular<sup>†</sup> diameters, the density is intermediate between that of bulk liquid and vapour. The number of atoms within the droplet, and the mean free energy of the droplet per atom, are thus only approximate concepts, which cannot be rigidly defined for any choice of reference interface.

The classical theory of nucleation ignores these difficulties. A liquid droplet is regarded as a sphere having the density of bulk liquid right up to some limiting geometrical surface, and the interfacial energy  $\sigma$  is defined with reference to this surface. Tolman (1949), Kirkwood and Buff (1949) and others have shown that  $\sigma$  expressed in this way will not be constant but will decrease with decreasing drop size. The magnitude of the effect depends on the nature of the interatomic forces, and it is very small when only nearest neighbour interactions are considered. Buff (1951) has attempted to modify eqn. (46.2) for the free energy change given below to allow for the variation of  $\sigma$ ; his theory suggests that values of  $\sigma$  calculated from experimental results using the classical nucleation theory will be substantially smaller than  $\sigma_\infty$ . Assuming Lennard-Jones atomic forces, however, Benson and Shuttleworth (1951) have shown that for the extreme case of a "droplet" consisting of a close-packed cluster of thirteen atoms,  $\sigma$  is only 15% less than  $\sigma_\infty$ .

Attempts at exact treatments which avoid the above inconsistencies have been made (see, for example, Reiss, 1952), but the statistical summations which replace the macroscopic parameters cannot usually be evaluated. Similar difficulties occur in some of the statistical theories of the liquid state (see Section 21). We shall accept the assumptions of the classical theory and shall not attempt to describe the more exact formulations of the problem. The variation of  $\sigma$  with droplet size, and the lack of precision in defining the size, are less serious than first appears, since the rate of nucleation is found to be determined principally by the properties of droplets in a restricted critical size range. Provided no formal attempt is made to identify  $\sigma$  with  $\sigma_\infty$  (they are sometimes found *empirically* to be almost equal), the difficulties which we have described in detail do not seem to invalidate the results. However, in recent years it has become apparent that there are other difficulties in the quantitative formulation of the theory of nucleation which cannot be entirely avoided by

<sup>†</sup> Strictly, we should use the more general word "molecule", but it seems rather artificial to write of metallic molecules since almost all metallic vapours are monatomic. For the remainder of the chapter we write only of atoms, but the discussion in this and other chapters often applies to non-metals if "molecule" is substituted for "atom".

the macroscopic model and which may be attributed, at least partially, to our lack of knowledge of the detailed atomic configurations in the liquid state, and especially of the atomic motions which contribute to its thermal entropy. These difficulties have resulted in some confusion about the correct expression for the equilibrium distribution of very small liquid droplets in a vapour, and are discussed below.

Suppose then that the number of atoms in a liquid droplet and in the vapour phase are  $n$  and  $N^v$  respectively, and that the chemical potentials, or free energies per atom, in the bulk phases are  $g^l$  and  $g^v$ . The  $n$  atoms which have condensed will adopt a spherical configuration so as to minimize the excess surface energy, and the free energy of the assembly will thus change by an amount  $\Delta G$  when the droplet is formed, where according to the classical theory

$$\Delta G = \Delta G' \quad (46.1)$$

$$\text{and} \quad \Delta G' = (4\pi r^3/3v^l)(g^l - g^v) + 4\pi r^2\sigma. \quad (46.2)$$

In this equation,  $r$  is the radius of the droplet, and  $v^l$  the volume per atom in the liquid state. It is often more convenient to focus attention on the number of atoms in the droplet, rather than on its radius, so the equation may be written in the alternative forms

$$\Delta G' = n(g^l - g^v) + O_n\sigma = n(g^l - g^v) + \eta n^{2/3}\sigma, \quad (46.3)$$

where  $O_n$  is the surface area of the droplet containing  $n$  atoms, and  $\eta$  is a shape factor. In the vapour-liquid transformation,  $\eta = O_n/n^{2/3} = (36\pi)^{1/3}(v^l)^{2/3}$ , but the same form of equation applies to other nucleation problems in which the nucleus is not spherical in shape. In solid transformations, the surface free energy is not isotropic, and a polyhedral nucleus may be formed. The change in free energy should then include a term in  $n^{1/3}$  representing the edge free energy; if this is neglected, eqn. (46.3) can still be used, giving appropriate mean values to  $\eta$  and  $\sigma$ . The use of this equation implies, of course, that only surface energies need be considered.

We have now to consider what are the permitted values of  $n$ . The association of  $n$  vapour atoms to form a cluster can clearly be regarded as a liquid droplet with less and less plausibility as  $n$  decreases, and it makes the theory a little less artificial if we impose a restriction that a liquid droplet must contain a certain minimum number  $p$  of atoms. The value of  $p$  is, in fact, immaterial to the final result.

In the classical theory of nucleation, eqns. (46.2) and (46.3) are regarded as giving the free energy change of the whole assembly when a liquid droplet is formed from the vapour, and this is equivalent to the assumption that if a bulk liquid is separated into a vapour of droplets at the same temperature and pressure, the free energy change per droplet is given by the surface energy terms in these equations. Frenkel (1939, 1946) pointed out, however, that a liquid droplet in a vapour should be regarded as a kind of macromolecule, and a number of authors (Rodebush, 1952; Kuhrt, 1952; Lothe and Pound, 1962) independently recognized that this might appreciably affect the magnitude of  $\Delta G$ . Any difference between  $\Delta G$  and  $\Delta G'$  essentially arises because a liquid droplet in a vapour differs from an equivalent region of bulk liquid not only in having a surface, but also in possessing much more freedom

to translate and rotate independently of the other droplets. Thus, according to Lothe and Pound, a more correct expression for the free energy change would be

$$\Delta G = \Delta G' + \Delta G_{\text{tr}} + \Delta G_{\text{rot}} + \Delta G_{\text{rep}}, \quad (46.4)$$

where  $\Delta G_{\text{tr}}$  and  $\Delta G_{\text{rot}}$  are contributions to the energy which may be considered to arise from the "heating" of a stationary droplet until it acquires the kinetic energy appropriate to the state defined by the temperature  $T$  and pressure  $p$  of the vapour. The term  $\Delta G_{\text{rep}}$  is the so-called replacement energy and arises because when the droplet acquires the gas-like translational and rotational motion it must lose six internal or liquid-like degrees of freedom. It is often stated that  $\Delta G'$  represents the free energy of formation of a stationary droplet, but this is misleading; for example, its centre of mass is not stationary. A more correct statement might be that  $\Delta G'$  represents the energy of a droplet confined as in a liquid rather than as in a vapour.

The terms  $\Delta G_{\text{tr}}$  and  $\Delta G_{\text{rot}}$  are both negative and may be calculated from the quantum-statistical contributions to the additional entropy of a droplet in the vapour, whilst  $\Delta G_{\text{rep}}$  is correspondingly positive since it represents a loss of entropy. A considerable controversy exists over the magnitude of  $\Delta G_{\text{rep}}$ , which at least partially reflects our ignorance of atomic motions in the liquid state. It is clear, for example, that if groups of atoms execute cooperative rotations in the liquid which are only slightly hindered, then one part of  $\Delta G_{\text{rep}}$  must almost cancel  $\Delta G_{\text{rot}}$ . Intuitively, it seems unlikely that the rotational component is cancelled in this way, but it is difficult to prove that it is not.

Since all the terms in  $\Delta G$  except the first terms in (46.2) or (46.3) represent the difference in free energy between a "gas" of liquid droplets and the same liquid in bulk form, it would be possible in principle to define  $\sigma$  in (46.2) or (46.3) so as to include the additional terms of (46.4). This would have the advantage of retaining the classical formulation of the equations of nucleation theory, but would involve the unsatisfactory feature of a surface free energy  $\sigma$  dependent on the size of the assembly. An alternative is to write explicit expressions for the remaining terms of (46.4) and hope that  $\sigma$  then corresponds approximately to the macroscopic free energy, despite the difficulties, which have already been stressed, of using macroscopic concepts for very small regions. This latter course has generally been followed in recent work. Note that when the size of the droplet becomes very large, the difference between (46.1) and (46.4) tends to zero, since both terms in (46.3) increase indefinitely with the size of the embryo  $n$ , whereas the additional terms in (46.4) are nearly independent of  $n$ .

The translational partition function for the three-dimensional motion of an embryo of size  $n$  in the standard state "gas" of such embryos (cf. eqn. (13.2)) is

$$Q_{\text{tr}} = (2\pi mkT/h^2)^{3/2} n^{3/2} v^v, \quad (46.5)$$

where  $m$  is the mass of an individual atom and  $v^v = kT/p$  is the molecular volume of the embryo. Clearly  $v^v = V/N^v$  is also the volume per atom in the vapour phase.

The rotational partition function is similarly

$$Q_{\text{rot}} = \pi^{1/2} (8\pi^2 kTI)^{3/2} / h^3, \quad (46.6)$$

where  $I$  is the moment of inertia of the spherical droplet and no atomic symmetry (i.e. symmetry number = 1) is assumed. Certain difficulties clearly arise in the concept of rota-

tion applied to a liquid droplet, but a detailed analysis by Nishioka *et al.* (1971) seems to show that the above formula is correct. For a special nucleus of radius  $r$ ,  $I = (\frac{2}{5})\pi nr^2$ , and eqn. (46.6) can be put in the form

$$Q_{\text{rot}} = C_8(2\pi mkT/h^2)^{3/2} v^l n^{5/2}, \quad (46.7)$$

where  $C_8 = (\frac{3}{4})(\frac{8}{5})^{3/2}\pi \simeq 4.8$ , and  $v^l$  is the volume per atom in the liquid phase. Hence

$$Q_{\text{tr}}Q_{\text{rot}} = C_8(2\pi mkT/h^2)^3 v^v v^l n^4, \quad (46.8)$$

and if we write

$$\Delta G_{\text{tr}} + \Delta G_{\text{rot}} + \Delta G_{\text{rep}} = kT \ln(Q_{\text{tr}}Q_{\text{rot}}/Q_{\text{rep}}) \quad (46.9)$$

it remains only to evaluate  $Q_{\text{rep}}$ .

Frenkel assumed that  $Q_{\text{rep}}$  is small and may be neglected, but subsequent estimates have ranged from small values ( $\sim 10^3$ ) to large values almost sufficient to cancel  $Q_{\text{tr}}Q_{\text{rot}}$ . Lothe and Pound originally considered the replacement partition function to be the entropy portion of the free energy of an atom in a liquid, and an argument for this in terms of a series of virtual processes was given by Feder *et al.* (1966). On the other hand, Kuhrt (1952), Dunning (1965, 1969), and others included the binding energy of the liquid atom in their estimate of  $Q_{\text{rep}}$ , and so obtained a much higher value. The Lothe-Pound expression is

$$Q_{\text{rep}}^l \simeq (2\pi n)^{1/2} \exp(s^l/k), \quad (46.10)$$

where  $s^l$  is the entropy per atom in the liquid. In later work, Lothe and Pound (1966) argued that  $Q_{\text{rep}}$  must correspond to six degrees of freedom for which the relative positions of the atoms in the region of bulk phase considered do not change. In the case of a solid, this corresponds to translational and torsional vibrations with

$$Q_{\text{rep}}^s = (kT/h\nu_D)^6, \quad (46.11)$$

where  $\nu_D$  is the Debye frequency. The authors suggest, moreover, that  $Q_{\text{rep}}^l \simeq Q_{\text{rep}}^s$ . More recently, Nishioka *et al.* (1971) have calculated  $Q_{\text{rep}}^s$  for a small solid crystallite by a numerical normal mode analysis and obtained a result appreciably larger than that given by (46.11). The typical numerical value given by (46.10) or (46.11) is  $\sim 10^2$ – $10^4$ , whereas the normal mode analysis gives  $Q_{\text{rep}}^s \sim 10^8$ . However, this discrepancy is stated to be largely due to the difference between the vibrational surface free energy of the bulk plane surface and that of the crystallite, which is included in eqn. (46.11) but excluded from the normal mode estimate.

If the vapour is supersaturated, the first of the two terms in (46.2) is negative, and the second is positive. Since these terms are proportional to  $r^3$  and  $r^2$  respectively, it follows that the influence of the second term will become less as  $r$  increases. The curve of  $\Delta G$  against  $r$  (or  $n$ ) will thus increase to a maximum and then decrease again. The position of the maximum is given by  $\partial\Delta G'/\partial r = 0$ , which leads to the critical radius of the droplet  $r_c$  as

$$r_c = 2\sigma v^l/(g^v - g^l). \quad (46.12)$$

The number of atoms in a drop of critical size is similarly

$$n_c = \left\{ \frac{2\eta\sigma}{3(g^v - g^l)} \right\}^3 = \frac{32\pi\sigma^3(v^l)^2}{3(g^v - g^l)^3}. \quad (46.13)$$

A nucleus of radius  $r_c$  is in unstable equilibrium<sup>†</sup> with the vapour. Droplets of radius  $r < r_c$  will tend to evaporate, since an increase in size leads to an increase in  $\Delta G$ , whilst droplets of radius  $r > r_c$  will tend to grow, since an increase in radius then decreases  $\Delta G$ . Droplets with  $r < r_c$  are often referred to as embryos, and those with  $r > r_c$  as nuclei. The vapour is unsaturated for embryos, and supersaturated for nuclei.

By treating the vapour as an ideal gas, we can find a relation between  $p_r$ , the vapour pressure with which a drop of radius  $r$  is in equilibrium, and  $p_\infty$ , the vapour pressure in equilibrium with a flat liquid surface. When the vapour pressure is  $p_\infty$ ,  $g^v = g^l$ , and from eqn. (46.12) when the vapour pressure is  $p_r$ ,  $g^v - g^l = 2\sigma v^l/r$ . Differentiating this expression

$$(v^v - v^l) dp = -(2\sigma v^l/r^2) dr$$

since  $dg^v = v^v dp$  and  $dg^l = v^l dp$  at constant temperature,  $v^v$  and  $v^l$  being the volumes per atom in the vapour and liquid states. We neglect  $v^l$  in comparison with  $v^v$ , and write  $v^v = kT/p$ . Integrating from  $p_\infty$  to  $p_r$ , while  $r$  varies from  $\infty$  to  $r$ , we find

$$\left. \begin{aligned} kT \ln(p_r/p_\infty) &= 2\sigma v^l/r \\ \ln i = \ln(p_r/p_\infty) &= 2\sigma v^l/rkT. \end{aligned} \right\} \quad (46.14)$$

In this equation, the ratio  $i = p_r/p_\infty$  is used to measure the supersaturation. A quantity more often used is the degree of supersaturation

$$j = (p_r - p_\infty)/p_\infty = i - 1.$$

The degree of supersaturation is often expressed as a percentage  $100j$ , instead of a fraction. For small supersaturations,  $j = i - 1 \approx \ln i$ .

When the modified eqn. (46.4) is used for the free energy change, eqns. (46.13) and (46.14) become

$$n_c = \left[ \left( \frac{2}{3} \eta \sigma \{g^v - g^l - (4kT/n_c)\} \right) \right]^3, \quad (46.15)$$

$$\ln(p_r/p_\infty) = (2\sigma v^2/rkT) - 4/n, \quad (46.16)$$

respectively, where the  $4/n$  factor comes from (46.8). With a typical value of  $n_c \sim 100$ , the corrected  $n_c$  is about 10% smaller than the "classical" value.

Equation (46.14) is another form of the Gibbs-Thomson or Thomson-Freundlich equation (p. 182-4), and eqn. (46.16) shows that it remains valid to a very good approximation in the modified theory. It follows that the maximum value of the free energy increase  $\Delta G_c$  is given sufficiently accurately by

$$\begin{aligned} \Delta G_c &= \Delta G'_c + \Delta G_{tr} + \Delta G_{rot} + \Delta G_{rep} \\ &= \Delta G'_c + kT \ln(Q_{tr}Q_{rot}/Q_{rep}), \end{aligned} \quad (46.17)$$

where  $\Delta G'_c$  is obtained by substituting (46.12) into (46.2) to give

$$\Delta G'_c = 4\pi\sigma r_c^2/3 \quad (46.18)$$

<sup>†</sup> The fact that the equilibrium is unstable, which is physically obvious, is expressed mathematically by the condition that  $\partial\Delta G/\partial r = 0$  gives a maximum ( $\partial^2\Delta G/\partial r^2 < 0$ ), whereas for stable or metastable equilibrium  $\partial\Delta G/\partial r = 0$  gives a minimum of  $\Delta G$ .

so that the excess free energy is equal to one-third of the surface free energy of the critically sized nucleus. We may also write this expression in the equivalent forms

$$\Delta G'_c = \frac{16\pi(v^l)^2 \sigma^3}{3(g^v - g^l)^2} = \frac{16\pi(v^l)^2 \sigma^3}{3k^2 T^2 (\ln i)^2}, \quad (46.19)$$

$$\Delta G'_c = \sigma \eta n_c^{2/3} / 3 = 4\sigma^3 \eta^3 / 27(g^v - g^l)^2. \quad (46.20)$$

If  $g^l > g^v$ ,  $\Delta G$  does not have a maximum value, but increases rapidly with  $r$ . Under these conditions, the vapour is stable, and any liquid embryo which forms will quickly evaporate again. The relation between  $\Delta G$  and  $r$  for an unsaturated and a supersaturated vapour is shown in Fig. 10.1.

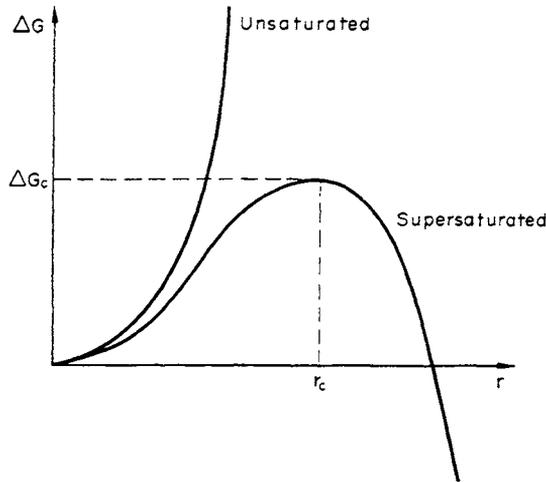


FIG. 10.1. Free energy of formation of a spherical liquid nucleus in a supersaturated and unsaturated vapour.

The above discussion assumes that liquid droplets form spontaneously in the interior of the vapour. In practice, it is well known that condensation usually occurs first on particles of dust or other impurities which may be present, or on the walls of the containing vessel. The impurities act by reducing the free energy barrier  $\Delta G_c$ , and nucleation under these conditions is said to be heterogeneous in contrast to the homogeneous nucleation discussed above. Homogeneous nucleation can only be obtained if special precautions are taken, but we shall first develop the theory for this process and consider later the modifications necessary when nucleation is heterogeneous.

#### 47. HETEROPHASE FLUCTUATIONS: VOLMER'S THEORY OF NUCLEATION

We see from the preceding section that the condensation of a supersaturated vapour requires the formation of nuclei of size  $r > r_c$ . The first satisfactory theory of nucleation was given by Volmer and Weber (1926) who assumed that there exists, effectively, a sta-

tionary distribution of embryos of size  $r < r_c$ . Frenkel (1939) generalized this theory to unsaturated vapour, and to other phase transformations, and we shall follow his treatment.

We have already seen that in a homogeneous phase there are local fluctuations of density, and of concentration if two or more components are present. These small fluctuations generally occur within the original phase, and their existence emphasizes the danger of applying the term homogeneous to very fine scale phenomena in macroscopic assemblies. In addition, we must consider the possibility of fluctuations which lead to local transitory phase transformations, so that small volumes can no longer be considered part of the original phase but have the atomic arrangement associated with a new phase. Fluctuations of this kind are called heterophase, and in a sense can be regarded as arising from a large homophase fluctuation. For example, a large density fluctuation in a small volume of vapour might lead to formation of a liquid droplet, or a large concentration fluctuation in a solid solution might lead to a rearrangement of atomic positions. Both of these are examples of heterophase fluctuations.

Heterophase fluctuations are, of course, responsible for the nucleation of phase transformations. Frenkel emphasized, however, that we must also regard them as existing in stable phases, the only difference being in the statistical distribution of the embryos. In a stable phase, the energy of an embryo of another phase (i.e. of a heterophase fluctuation) increases rapidly with its size, and the number of embryos present in equilibrium thus decreases extremely rapidly with size. For a metastable phase, the energy only increases initially, and then decreases again. If we attempt to find an equilibrium distribution of embryos in a metastable phase, we find that all the atoms are taking part in very large heterophase fluctuations; this is merely another way of expressing the condition that another phase has lower free energy. In nucleation problems of interest, we require the distribution of embryos in the initial stages of transformation, i.e. subject to the restrictive condition that almost all the atoms are present in the metastable state. The simplest assumption is that the distribution of embryos of size  $n < n_c$  is the same as the equilibrium distribution if the phase were in fact stable. This is the basis of the Volmer theory of nucleation.

Fluctuations of density or concentration produce changes in the free energy. If a fluctuation of any kind produces a rise in the free energy  $\Delta G$ , the probability of its occurrence in an equilibrium state of the assembly is proportional to  $\exp(-\Delta G/kT)$ . The number of embryos of size  $n$  is thus given by

$$N_n = N^v \exp(-\Delta G_n/kT), \quad (47.1)$$

where  $N_n$  is the statistical distribution function for embryos containing  $n$  atoms,  $N^v$  is the number of atoms in the vapour phase, and  $\Delta G_n$  is the standard free energy change resulting from the conversion of vapour into embryos. We shall suppose that the assembly may be characterized by the distribution function  $N_n$  for the most probable type of embryo, so that  $\Delta G_n$  is a function only of  $n$ . Strictly, this is a simplification of the problem, since we should also specify the shape of the embryos, but there will be a negligible error in limiting consideration to the shape corresponding to minimum surface energy for a given size. The procedure is essentially equivalent to replacing a summation by its largest term, which often has to be used in statistical mechanics. It then follows that  $\Delta G_n$  may be identified with the

free energy change of eqn. (46.4), and that the total number of atoms in the assembly is given by

$$N = N^v + \sum N_n n. \quad (47.2)$$

In the classical theory of nucleation, as developed in most books, the equilibrium distribution is represented by an equation similar to (47.1) but with  $\Delta G_n$  replaced by  $\Delta G'_n$  from eqn. (46.2). Feder *et al.* (1966) point out that if this were valid, the concentration  $N_n/V$  of embryos of size  $n$  in a fixed volume  $V$  would vary as  $p^{n+1}$ , which is contrary to the law of mass action. (One power of  $p$  comes from the pre-exponential factor ( $N^v/V$ ) and the others from the exponential term, since  $n(g^v - g^l) = nkT \ln i$ .) It follows from this, as well as from the preceding discussion, that eqn. (47.1) cannot be correct unless the additional terms of (46.4) are included in  $\Delta G_n$ . However, since these additional terms are almost independent of  $n$ , it is convenient to bring them into the pre-exponential and thus to write

$$N_n = N^v(Q_{tr}Q_{rot}/Q_{rep}) \exp(-\Delta G'_n/kT), \quad (47.3)$$

where the quantity in the first bracket represents the Lothe-Pound correction factor. In the form (47.3), the concentration ( $N_n/V$ ) varies correctly as  $p^n$  (since one power of  $p$  has been cancelled by the factor  $v^v = V/N^v$  in the translational partition function (eqn. (46.5)).

Frenkel gave a statistical-mechanical derivation of eqn. (47.1) for the case of homophase fluctuations, which was reproduced in the first edition of this book, but is here omitted for brevity. However, it is apparent that we have arrived at the equilibrium distribution by a rather tortuous path involving both thermodynamics and statistical mechanics, and it is possible to use the more direct procedure of minimizing the free energy of a vapour containing embryos with respect to the concentration of embryos of given class. This gives directly

$$N_n = Q_n \exp(ng_v/kT), \quad (47.4)$$

where  $Q_n$  is the partition function for an embryo of size  $n$  in the vapour phase. Reiss and Katz (1967) and Reiss *et al.* (1968) have attempted to evaluate  $Q_n$  using a method known as the standard phase integral, and they obtain a result

$$\begin{aligned} N_n &= \frac{Q_{tr}^c}{Q_{tr}^d} \exp(-\Delta G'/kT) \\ &= (V/v^d) \exp(-\Delta G'/kT), \end{aligned} \quad (47.5)$$

where  $Q_{tr}^c$  is the translational partition function for a liquid cluster in the vapour phase and  $Q_{tr}^d$  is the corresponding partition function for a droplet in the bulk liquid. For  $Q_{tr}^c$ , the volume  $v^v$  of eqn. (46.5) is replaced by the volume of the whole assembly  $V$ , whilst for  $Q_{tr}^d$  it is replaced by the volume  $v^d$  in the bulk liquid over which the centre of mass of a droplet with fixed boundaries fluctuates. (In their papers a distinction is made between a cluster, which is a small liquid region in the vapour, and a droplet, which is an equivalent region in the bulk liquid.) Essentially the same result is obtained by Lin (1968) using the method of

the grand canonical ensemble. Thus the correction to the classical distribution is given by Lothe and Pound as

$$\Gamma_{LP} = Q_{tr}Q_{rot}/Q_{rep} \tag{47.6}$$

and by Reiss *et al.* as

$$\Gamma_R = V/N^v v_d = v^v/v^d. \tag{47.7}$$

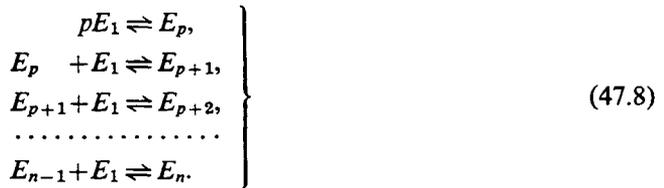
The rotational contribution does not appear in the Reiss equation, and according to Nishioka *et al.* (1971) this is because the authors suppose, incorrectly, that the effect of surroundings on a drop in bulk liquid may be completely simulated by enclosing the atoms in a fixed volume. Apart from the rotational contribution, however, the two equations should be equivalent, and the difficulty of calculating  $Q_{rep}$  has been replaced by that of calculating  $v^d$ . Reiss *et al.* believe that the correction (47.6) as used by Lothe and Pound is partially redundant because some states are being counted twice; Nishioka *et al.*, on the other hand, believe that  $v^d$  is much smaller than the volume used by Reiss *et al.*

We have emphasized these difficulties because the numbers involved are very large. For a critical nucleus size  $n_c \sim 100$ , applicable to water vapour for example, Lothe and Pound's original estimate for the factor (47.6) was  $10^{17}$ , whilst Reiss *et al.* estimate from (47.7) a factor of  $10^3-10^6$ . A rough estimate of  $v^d$  due to Nishioka *et al.* is

$$v^d = (8v^{1/3})(3n+1)^{-2},$$

and this gives typically  $\sim 5 \times 10^8$  for the factor in (47.7). Thus the differences amount to a factor of  $10^2-10^5$  in the estimate of the translational contribution, and a disagreement about whether or not there is a rotational contribution.

Volmer assumed that an embryo is formed as a result of a large number of small scale (bimolecular) fluctuations, rather than by a sudden large fluctuation. If we use the symbol  $E_n$  to denote an embryo containing  $n$  atoms, and  $E_1$  to denote a single atom, the process of formation may be written



The symbol  $E_p$  is used for the droplet containing  $p$  atoms, which we decide to regard as the smallest recognizable heterophase fluctuation (see p. 424). An embryo is thus assumed to grow or shrink by the addition or removal of individual atoms. When the primary phase is a vapour, the forward reactions of the above set will result from collisions between embryos and vapour atoms; the probability of embryo-embryo collisions is clearly negligible.

Consider now an unsaturated vapour ( $i < 1$ ), in which the free energy of an embryo increases continually and rapidly with  $n$ . A state of dynamic equilibrium will be attained, in which the number of embryos of given size will remain effectively constant, although individual embryos are constantly growing or evaporating. Any non-stationary distribution will rapidly change towards the equilibrium distribution. Suppose, for example, that

embryos  $E_n$  are being formed more rapidly than they disappear into either  $E_{n-1}$  or  $E_{n+1}$ . The concentration of  $E_n$  embryos will then increase with time, and this will increase the number decomposing until equilibrium is attained. The relaxation time for the attainment of the stationary distribution depends on kinetic features of the growth of embryos; it is possible that this has some importance in solid-state reactions (see pp. 446-52), but we shall neglect it in our discussion of the vapour-liquid change, since it will certainly be extremely small.

The fundamental condition for the equilibrium state can be expressed mathematically as follows. There will be a certain probability per unit time that an atom will condense on the surface of a  $E_n$  embryo, converting it into a  $E_{n+1}$  embryo. If each vapour atom which collides with the embryo condenses, the probability is given approximately by the product of its area and the kinetic theory collision factor  $p/(2\pi mkT)^{1/2}$ , where  $p$  is the pressure and  $m$  the atomic mass. More generally, however, we write this probability as  $q_0$  per unit area per unit time, in order to preserve the same symbol for condensed state transformations. For vapour-liquid transformations,  $q_0$  is given by the product of the collision factor and a condensation or accommodation coefficient  $\alpha_c \ll 1$ , and it is usual to assume that  $q_0$  is independent of  $n$ . Actually, there may be a rather rapid variation of  $\alpha_c$ , and hence of  $q_0$ , with  $n$  for very small embryos, but, as we shall show, this probably does not affect the final result. There will also be a probability  $q_n$  per unit area per unit time that an atom will evaporate from the surface of an embryo of  $n$  atoms converting it into a  $E_{n-1}$  embryo. In this case,  $q_n$  will not be independent of  $n$ , since the probability of evaporation is a function of the free energy of the embryo. Now consider the net transfer of atoms between embryos of size,  $n, n+1$ , as represented by the equations

$$\left. \begin{aligned} E_n + E_1 &= E_{n+1}, \\ E_{n+1} - E_1 &= E_n. \end{aligned} \right\} \quad (47.9)$$

The rate of the first process is  $N_n O_n q_0$  and of the reverse process is  $N_{n+1} O_{n+1} q_{n+1}$ . In unit time, therefore, the net transfer is given by

$$I = N_n O_n q_0 - N_{n+1} O_{n+1} q_{n+1} = 0. \quad (47.10)$$

The expression is equated to zero, since the numbers,  $N_n, N_{n+1}$  are assumed to be the numbers of embryos in the equilibrium distribution. It then follows from the principle of detailed balancing (p. 83) that the net transfer of atoms in any separate process of the kind considered must be zero. Equation (47.10) is the fundamental expression of the equilibrium state; the corresponding equation for the metastable state (when  $I \neq 0$ ) is the basis of the Becker-Döring theory of nucleation.<sup>†</sup>

<sup>†</sup> It should be remarked that in developing the Becker-Döring theory in his book *Kinetik der Phasenbildung*, Volmer gives this equation in different form. He assumes that a vapour atom effectively collides with an embryo when its centre passes through the surface of a sphere having radius greater by an atomic radius than the radius of the embryo itself. He writes the area of this effective collision sphere as  $O'_n$ , and he further assumes that an atom escapes from the  $n+1$  embryo when its centre passes through the same limiting surface. Thus the quantities  $O_n, O_{n+1}$  in eqn. (47.10) are both replaced by  $O'_n$ . In view of the general inapplicability of macroscopic concepts to the embryos, it seems very doubtful whether refinements of this kind are worth while.

Now consider a supersaturated vapour ( $i > 1$ ). Equation (47.1) then gives the result, as anticipated on p. 429, that  $N_n$  tends to infinity as  $n$  tends to infinity; all the atoms are thus present as large embryos, and there is no vapour phase. We are interested, however, in the embryo distribution in the initial stages of the transformation, when almost all the assembly is in the vapour phase, and the most probable states are thus excluded. Volmer's assumption is that nuclei of size greater than  $n_c$  grow rapidly, and may be regarded as removed from the assembly so far as the calculation of nucleation rates is concerned. For embryos of size smaller than  $n_c$ , the distribution function is assumed to be given by eqn. (47.1). We shall use the symbol  $Z_n$  for the number of embryos of size  $n$  present in a metastable assembly, making a distinction between this distribution function and  $N_n$  which applies to a stable assembly. The original Volmer theory thus requires

$$\left. \begin{array}{l} n < n_c \quad Z_n = N \exp(-\Delta G_n/kT), \\ n > n_c \quad Z_n = 0. \end{array} \right\} \quad (47.11)$$

This distribution can be obtained formally by erecting an infinite free energy barrier at  $n = n_c$ , but there would not then, of course, be any nucleation. To allow nucleation, we must suppose that each embryo of size  $n_c$  can penetrate the barrier when a vapour atom condenses on it, and it then becomes a nucleus which is removed from the assembly. The distribution of embryos will be maintained if  $n_c + 1$  vapour atoms are added to the assembly for each nucleus removed. This rather unreal situation is sometimes called a quasi-steady-state distribution. The rate of nucleation will be given by the product of the number of critical size embryos in the quasi-steady-state,  $Z_c$ , and the probability in unit time of a vapour atom condensing on one of these embryos.

The preceding paragraph gives the mathematical conditions required to ensure a quasi-steady-state distribution. Physically, there will clearly be a close approximation to such a distribution at the beginning of transformation, when the supply of vapour atoms is virtually infinite, provided the rate of nucleation is small. Steady-state nucleation is illustrated diagrammatically in Fig. 10.2, the upper part of which shows the curve of  $\Delta G_n/kT$  against  $n$ . The horizontal lines are drawn with a weight proportional to  $\exp(-\Delta G_n/kT)$ , so that they illustrate the relative numbers of embryos which would be present if  $\Delta G_n$  continued to increase beyond  $n = n_c$ . Nuclei of size  $n_c$  just spill over the free energy barrier and, in the steady state, the rate at which they do so will be equal to the rate at which new embryos of size  $n_c$  are formed. The lower part of the figure shows the distributions given by eqn. (47.11) (full line) and eqn. (47.1) (dotted line).

From (47.11) we can now write the number of stable nuclei formed in the assembly in unit time as

$$I = q_0 O_c Z_c = N q_0 O_c \exp(-\Delta G_c/kT). \quad (47.12)$$

These nuclei are formed by a chain of the forward processes of (47.8). It should be noted that most embryos evaporate again before becoming nuclei; only occasionally will there be a series of favourable fluctuations. The quantity  $I$ , called the nucleation rate or nucleation current, is of great importance. It has dimensions  $s^{-1}$ , and is proportional to the total

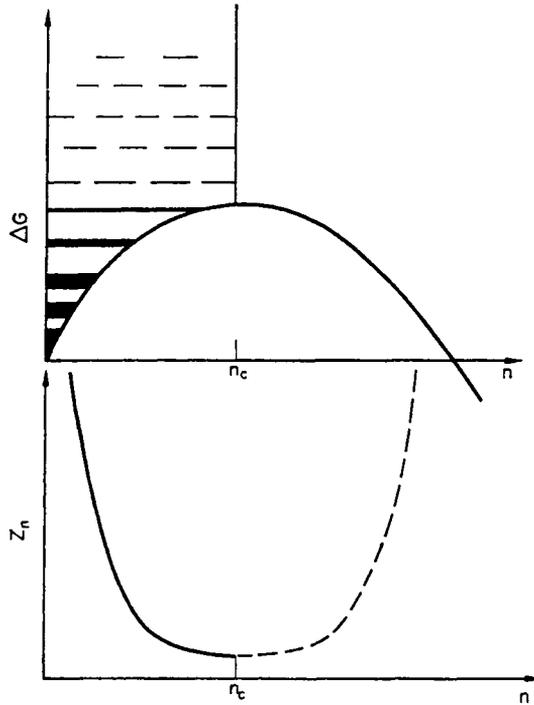


FIG. 10.2. To illustrate the quasi-steady distribution of embryos of size  $n < n_c$  (after Fisher *et al.*, 1948). The upper figure shows the curve of free energy of formation vs. nucleus size, and the horizontal lines are drawn to indicate qualitatively the numbers of embryos of various sizes present in a quasi-steady distribution. The lower figure shows the distribution assumed in the Volmer theory (eq. (47.11)).

number of atoms in the assembly; this is true of all homogeneous nucleation phenomena. The number of nuclei formed in unit time in unit volume is thus constant, and is given by

$$vI = \frac{I}{Nv^j} = \frac{q_0 O_c}{v^j} \exp(-\Delta G_c/kT). \quad (47.13)$$

We emphasize again that this constant nucleation rate only applies to conditions under which  $I$  is small and little or no transformation has occurred. For the vapour-liquid change, and in some other transformations, the growth rate at constant supersaturation is much greater than the steady nucleation rate. The kinetics of the transformation are then effectively governed by the nucleation phenomenon, since a small number of nuclei effect the whole change. The transformation rate will be greater for large assemblies, and under specified conditions will be roughly proportional to the number of systems in the assembly. In the extreme case, the whole assembly will transform when a single nucleus has formed, and the reciprocal of  $I$  is the time required to do this. Since nucleation is a statistical process, however, the time to form one nucleus will show small variations in similar assemblies, or in the same assembly in different experiments.

From eqns. (46.19) and (47.12), the nucleation rate varies as

$$\exp(-1/(\ln i)^2) = \exp(-1/j^2),$$

and is thus extremely sensitive to the degree of supersaturation. The curve of  $I$  against  $i$  is shown in Fig. 10.3; the nucleation rate is very small until a critical supersaturation is reached, and then it rises abruptly to a much larger value. The critical supersaturation corresponds to the breakdown of the metastable state of the assembly, and the quasi-steady-state distribution of embryos can no longer be maintained.

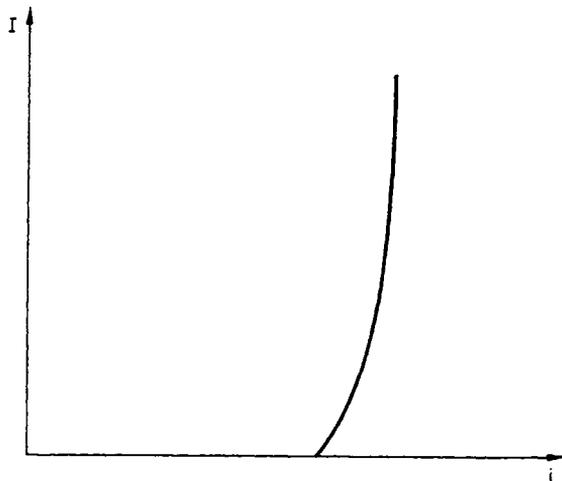


FIG. 10.3. Schematic variation of nucleation rate with degree of supersaturation.

As will be seen in the following sections, a value for a steady-state nucleation rate can be obtained for all kinds of transformations. The overall rate of transformation under conditions in which both the nucleation frequency and the growth rate have some influence was considered in Section 4, Chapter 1. The question then arises whether the empirical nucleation frequency used in the theory of Section 4 can be identified with the steady-state nucleation rate per unit volume  ${}^vI$ . The phenomenological theory uses the nucleation rate in the *untransformed* regions of the assembly, which we may treat as sub-assemblies where the conditions for quasi-steady-state embryo distribution are satisfied. Provided that the relaxation time for the establishment of the quasi-steady state is not too large, we may thus expect the nucleation rate of Section 4 to be the steady-state nucleation rate appropriate to the untransformed regions at any stage of the transformation. During the course of the transformation, however, the intensive parameters such as density, internal strains, composition, etc., which specify the condition of the untransformed part of the assembly, may be altered, producing corresponding changes in  ${}^vI$ . Thus although  ${}^vI$  of Section 4 may usually be identified with a steady nucleation rate, it cannot be assumed to be time independent, as has already been emphasized. Furthermore, if the assembly is suddenly changed (quenched)

from one condition to another at time  $t = 0$ , there may be a finite period during which the quasi-steady-state embryo distribution appropriate to the new condition is established. The initial nucleation rate will then vary rapidly with time.

#### 48. THE BECKER-DÖRING THEORY OF NUCLEATION†

Volmer and Weber's expression for the rate of nucleation consists of two terms—an exponential factor involving the free energy increase due to the formation of a critical size nucleus, and a factor which is proportional to the frequency of collision of vapour atoms. The main defect of the theory lies in the assumption that the steady state distribution function is given by (47.11). In the metastable assembly, critical size embryos can either grow or shrink with equal probability ( $q_n = q_0$  when  $n = n_c$ ), and nuclei of greater than critical size may also shrink again, although they are rather more likely to grow and thus be removed from the assembly. The true distribution function in the quasi-steady state will thus not fall abruptly to zero at  $n = n_c$ , but will decrease slowly, becoming effectively zero when  $n$  is large. This quasi-steady distribution is shown in Fig. 10.4; it approximates to (47.11) for very small  $n$  values, but progressively decreases below this distribution as  $n$  increases, the value of  $Z_c$  being half that given by (47.11).

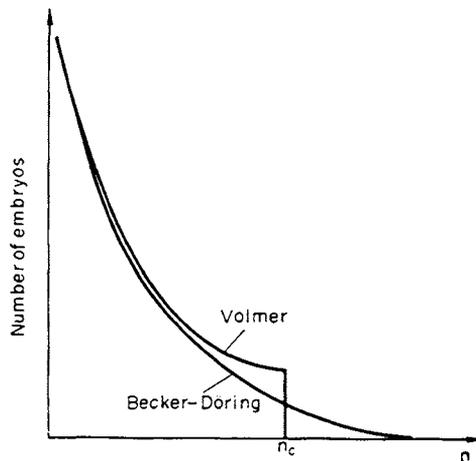


FIG. 10.4. Distribution functions for embryos of different sizes according to Volmer and Becker-Döring theories of nucleation.

Improvements of Volmer's equation were suggested by many workers, and notably by Becker and Döring (1935), who developed a kinetic theory of nucleation which is the basis of almost all subsequent treatments. The importance of this theory arises more from the correct formulation of the kinetic problem (within the limits discussed on pp. 422-3) than

† Much of the credit for the kinetic theory of nucleation properly belongs to Farkas (1927) who determined the so-called non-equilibrium or Zeldovich factor (see p. 439) well before the work of Becker and Döring. However, the theory is usually ascribed to the latter authors who developed it in a different way.

in the precise solution obtained; the result modifies only the non-exponential factor of eqn. (47.12), and the nucleation rate is much more sensitive to slight changes in  $\Delta G_c$  than to large changes in this factor. There is, moreover, no general agreement on the best value of the pre-exponential factor, and slightly different results are obtained in different developments of the Becker–Döring theory. This situation arises because the quasi-steady state is found to be characterized by a set of difference equations. An analytical solution can only be obtained by using approximations in which summations are replaced by integrals, and the results differ somewhat according to the procedure adopted. The derivation given here uses an approach due originally to Zeldovich (1943) and followed by Frenkel (1946), although the final result differs from that which Frenkel obtained.

Consider again the interchange of embryos  $E_n, E_{n+1}$  represented in eqn. (47.9). In place of (47.10), the net transfer is now given quite generally by

$$I_{n,t} = Z_{n,t}q_0O_n - Z_{n+1,t}q_{n+1}O_{n+1}, \tag{48.1}$$

where  $Z_{n,t}$  is the number of embryos of size  $n$  present in the assembly at time  $t$ . Using the result of (47.10), we may write this kinetic equation in the useful form

$$I_{n,t} = N_nq_0O_n \left\{ \frac{Z_{n,t}}{N_n} - \frac{Z_{n+1,t}}{N_{n+1}} \right\}, \tag{48.2}$$

and there will be a set of such equations for all allowable values of  $n$ . In general, the equations are very difficult to handle, but for the present we are interested only in the limiting conditions which lead to the quasi-steady-state distribution. As in the previous discussion, the required conditions are that  $I_{n,t}$  is small, and all the  $Z_{n,t}$  are small compared with the number of vapour atoms. The distribution function is then  $Z_{n,t} = Z_n$ , independent of time, and the steady state is maintained if  $I_{n,t} = I$ , a constant. The net rate,  $I$ , at which embryos  $E_{n-1}$  change to embryos  $E_n$  is equal to the rate at which  $E_n$  change to  $E_{n+1}$ , and so on.  $I$  is thus the rate at which nuclei are produced in the assembly.

We have no information about the details of the distribution function,  $Z_n$ , but we know that for very small embryos it must be effectively identical with the equilibrium distribution function which would be obtained in the absence of nucleation, whilst for large embryos, the distribution function must approach zero. Thus

$$n \rightarrow 0, \quad Z_n \rightarrow N_n; \quad n \rightarrow \infty, \quad Z_n \rightarrow 0.$$

It is convenient to suppose that  $Z_n = N_n$  for all values  $n \leq p$ , and  $Z_n = 0$  for all values  $n \geq s$ , where  $s$  is greater than  $n_c$ . The latter condition replaces Volmer's assumption that  $Z_n = 0$  for  $n = (n_c + 1)$ , and we shall see that it is not necessary to specify the values of either  $p$  or  $s$ .

From (48.2), we may now write a series of equations

$$\frac{I}{N_nq_0O_n} = \frac{Z_n}{N_n} - \frac{Z_{n+1}}{N_{n+1}}$$

for all values of  $n$  from  $p$  to  $s$ . Adding these equations together,

$$I \sum_p^s \frac{1}{N_n q_0 O_n} = \frac{Z_p}{N_p} - \frac{Z_{s+1}}{N_{s+1}} = 1$$

or

$$I = \frac{q_0}{\sum_p^s \left\{ \frac{1}{N_n O_n} \right\}} \quad (48.3)$$

In this expression,  $N_n$  is determined by eqns. (47.1) and (46.3), and it follows that the discontinuous function  $1/N_n$  has a sharp maximum at  $n = n_c$ .† The only effective contributions to the summation will come from terms where  $n - n_c$  is small, and the effect of terms with large or small  $n$  may be neglected. In all the important terms,  $O_n$  will thus differ only slightly from the surface area of the critical nucleus,  $O_c$ , and it is thus permissible to treat this as a constant factor and take it outside the summation giving

$$I \approx \frac{q_0 O_c}{\sum_p^s \frac{1}{N_n}} \quad (48.4)$$

In order to proceed further, we make the approximation of treating  $N_n$  as a continuous function of  $n$ , although strictly it is only defined for integral values. The summation can then be replaced by an integral. Furthermore, the function  $1/N_n$  has appreciable values only near  $n = n_c$ , so that it is permissible to expand  $\Delta G_n$  as

$$\Delta G_n \approx \Delta G_c + \frac{\xi^2}{2} \left( \frac{\partial^2 \Delta G_c}{\partial \xi^2} \right)_{\xi=0}, \quad (48.5)$$

where  $\xi = n - n_c$ . The linear term in  $\xi$  is, of course, zero, since  $(\partial \Delta G_n / \partial \xi)_{\xi=0} = 0$ ,  $\Delta G_n$  having its maximum value at  $n = n_c$ . Using eqns. (48.5), (47.1), (46.3), and (46.20), we may thus write

$$\sum_p^s \frac{1}{N_n} \approx \frac{1}{N} \exp\left(\frac{\Delta G_c}{kT}\right) \int_{p-n_c}^{s-n_c} \exp\left(\frac{-\Delta G_c \xi^2}{3n_c^2 kT}\right) d\xi \quad (48.6)$$

Finally, since only values of  $1/N_n$  near  $\xi = 0$  have any appreciable influence on the value of the integral, the limits  $\xi = -(n_c - p)$ ,  $\xi = s - n_c$  may be changed to  $\xi = \pm \infty$  without affecting the result. The integral is then transformed into the error integral and

$$\sum_s^p \frac{1}{N_n} \approx \frac{1}{N} \exp\left(\frac{\Delta G_c}{kT}\right) \cdot \left(\frac{3\pi kT}{\Delta G_c}\right)^{1/2} n_c \quad (48.7)$$

† All treatments of the Becker-Döring theory involve summations of sets of terms which have appreciable values only near  $n = n_c$ . It is interesting to note here that by taking the simplest and most drastic approximation — replacing the sum by its maximum term — the Volmer equation (47.12) for nucleation is obtained. This clearly cannot be correct, since the other terms near  $n = n_c$  will make finite contributions, and the true nucleation rate must be smaller than that given by (47.12).

Substituting into (48.4), the nucleation current is finally obtained as

$$I = \frac{Nq_0O_c}{n_c} \left( \frac{\Delta G_c}{3\pi kT} \right)^{1/2} \exp\left(\frac{-\Delta G_c}{kT}\right). \quad (48.8)$$

The factor

$$\Gamma_Z = (1/n_c)(\Delta G_c/3\pi kT)^{1/2} \quad (48.9)$$

by which eqn. (48.8) differs from the Volmer eqn. (47.13) is sometimes called the Zeldovich factor, although as we have already noted it was first derived by Farkas (1927). A typical value is  $\frac{1}{20}$  and it is unlikely to be smaller than  $10^{-2}$  in conditions under which experimental observations may be made. Because of the very rapid variation of nucleation rate with supercooling or supersaturation ratio (see Fig. 10.3), changes in the pre-exponential factors of this magnitude have very little effect on observable features of the kinetics.

The derivation above is identical with that used by Reiss (1952), and differs from Becker and Döring's original method, which was to write (48.1) in the form

$$IR_n = \varphi_n - \varphi_{n+1}$$

where

$$R_n = \frac{1}{q_0} \prod_2^n \left( \frac{q_n}{q_0} \right), \quad \varphi_n = Z_n O_n \prod_2^n \left( \frac{q_n}{q_0} \right).$$

Summing for all values of  $n$  from 1 to  $s$  ( $\varphi_1 = Z_1 O_1$ ,  $\varphi_s = 0$ )

$$I \sum_1^s R_n = Z_1 O_1. \quad (48.10)$$

The "growth resistances"  $R_n$  are evaluated by expressing the terms in the product as exponential functions of the reciprocal radii, using eqn. (46.14). This gives

$$R_n = \frac{1}{q_0} \exp\left\{ \frac{2\sigma v_L}{kT} \left( \sum_2^n \frac{1}{r_n} - \frac{n-1}{r_c} \right) \right\} \quad (48.11)$$

and the sum is approximated by an integral. The resistances  $R_n$  have a sharp maximum at  $n = n_c$ , and the sum  $\sum_1^s R_n$  may thus be transformed into the error integral by expanding about  $n = n_c$  and changing the limits of integration, as above.

Volmer (1939) followed the Becker–Döring treatment, but he pointed out that the sum in (48.11) had previously been overestimated by an amount representing the heat of solution of one molecule in a large volume of liquid. His corrected expression for the nucleation rate then contained an additional factor  $\exp(\Delta h^{lv}/kT)$ , where  $\Delta h^{lv}$  is the molecular heat of condensation. This factor is not negligible (it may be 5–7 powers of ten!), but Volmer states it is accidentally cancelled by another previously neglected effect. This is the sudden decrease in the condensation coefficient, and hence of  $q_0$ , at very low  $n$  values, due to the heat released by condensation. Both of Volmer's "neglected" factors seem to concern the properties of embryos with very small  $n$  values, and it is a weakness of Becker and Döring's original procedure that the initial summation has to be made from  $n = 1$  to  $n = s$ , instead

of from  $n = r$  to  $n = s$ . It is thus probable that the cancellation of the factors is not really accidental, since the treatment in terms of distribution functions shows how their introduction may be avoided. However, this argument only applies to the abnormally low values of the condensation coefficient for small embryos, and  $q_0$  should undoubtedly contain a condensation coefficient appropriate to critical-sized embryos in addition to the collision factor. It has often been assumed that the condensation coefficient is essentially unity, but under some circumstances it may be appreciably smaller, as emphasized by Hirth and Pound (1963).

The effect of the latent heat of condensation was considered more carefully by Kantrowitz (1951) who pointed out that the mean temperature of embryos of size  $n$  is slightly greater than the temperature of the vapour itself because of the energy released by a condensing atom. His quantitative estimate of the temperature rise is very nearly zero for small embryos, but is a few degrees Celsius for values of  $n$  near  $n_c$ ; the corresponding nucleation rate is changed by a factor which is a function of the appropriate heat content quantities. In a more detailed treatment of this problem, Feder *et al.* (1966) considered coupled currents of matter and thermal energy in a non-isothermal nucleation process. They found that the nucleation current of sub-critical embryos is largely carried by embryos which are colder than the vapour. In each size class of embryos there are more embryos which are at a temperature above that of the vapour than there are at a lower temperature, thus giving a mean temperature in excess of that of the vapour. However, the contribution to the nucleation rate of these more numerous "warm" embryos is more than counterbalanced by the higher probability of growth of the "cold" embryos. The net effect on the nucleation rate is to multiply by a factor

$$\Gamma_{\text{Th}} = \frac{(c'_v + \frac{1}{2}k)kT^2}{(c'_v + \frac{1}{2}k)kT^2 + (\Delta h^{lv})^2} \quad (48.12)$$

in which  $c'_v$  is the specific heat of the liquid at constant volume. (The expression is modified somewhat when nucleation is carried out in the presence of an inert carrier gas.) A typical value of  $\Gamma_{\text{Th}}$  is  $\sim \frac{1}{5}$ , so that it has a very small influence in the pre-exponential term.

The final expression for the nucleation rate may now be written in the equivalent forms

$$I = \Gamma_z \Gamma_{\text{Th}} N_c q_0 O_c \quad (48.13a)$$

$$= N \Gamma_z \Gamma_{\text{Th}} q_0 O_c \exp(-\Delta G_c/kT) \quad (48.13b)$$

$$= N \Gamma_z \Gamma_{\text{Th}} \Gamma_{\text{LP}} q_0 O_c \exp(-\Delta G'_c/kT). \quad (48.13c)$$

Accurate experimental measurements of  $I$  are not possible because of its rapid variation with supersaturation (or supercooling), and it is usually possible to determine only the critical supersaturation at which the nucleation rate changes from a negligible to a very large value. As already emphasized, this means that the results are not sensitive to the value assumed for the pre-exponential term to within a few powers of ten. This means that the Zeldovich factor and the non-isothermal factor may be effectively ignored, and to within a sufficient approximation

$$I = (\text{collision frequency of vapour atoms}) \exp(-\Delta G_c/kT). \quad (48.14)$$

However, the very large values calculated for  $I_{LP}$  by some authors certainly cannot be ignored in a comparison of theory and experiment. Becker and Döring applied their theory to the condensation of water vapour and obtained good agreement with the experimental results of Volmer and Flood (1934); the critical nucleus size was estimated at  $\sim 100$  molecules. Volmer and Flood also determined the critical supersaturation ratio for various organic vapours, and their results were compared with the theory by Hollomon and Turnbull (1953). The most convenient test is to work backwards from the observed supersaturation to calculate  $\sigma$  and then compare this with the macroscopic value  $\sigma_\infty$ . Very good agreement is found, and for many years this was taken to imply that the Volmer-Becker-Döring theory is an adequate description of the nucleation process, despite the qualifications on pp. 422-3. The results do not agree with the predictions of Buff's modified theory that  $\sigma \simeq 0.8\sigma_0$ .

If the Lothe-Pound theory is correct, the additional factor of  $10^{17}$  clearly cannot be neglected, and in fact this seems to destroy the above agreement between theory and experiment. More recent experimental work, however, has given results which seem to agree with the predictions of the Lothe-Pound theory. Jaeger *et al.* (1969) and Dawson *et al.* (1969) used a supersonic air nozzle technique, instead of a cloud chamber, to measure the critical supersaturation ratio for water, ammonia, benzene, ethanol, chloroform, and freon. Their results for water confirm the previous work, which agrees with the classical theory, and those for ethanol were inconclusive; in all other cases, much higher nucleation rates (i.e. lower critical supersaturations) than those predicted by the classical theory were found, and these were in reasonable agreement with the estimated Lothe-Pound factor. The authors point out that the classical theory seems to apply for polar or rod-shaped molecules, and Abraham (1969) has noted that liquids which are mainly hydrogen bonded agree with classical theory, and that such liquids have abnormally low surface entropies. This leads to an increase in the surface free energy of the less-ordered embryo over the flat surface of the bulk liquid, and the predicted change is of the required magnitude to restore agreement with (48.13c). In other words, Abraham suggests that the error in  $\Delta G'_c$  resulting from incorrect use of the macroscopic  $\sigma$  is just sufficient to cancel the other terms in (46.4).

#### 49. NUCLEATION OF THE SOLID FROM THE VAPOUR OR THE LIQUID

The theory developed above applies specifically only to the condensation of a supersaturated vapour. This is a transformation rarely studied in metallic assemblies, and the space devoted to it may therefore seem excessive. The results obtained, however, provide a basis for the treatment of other transformations, and we can now consider the relation of the theory to transformations involving solid phases.

If we consider first the growth of solid crystals from the vapour phase, we expect the nucleation rate to be given by an equation similar to (48.8) or (48.13). The embryo crystal of minimum surface energy, however, will not have a spherical form, but a shape conforming to Wulff's construction, described in Chapter 5. This modification to the form of the surface energy term is not the only alteration needed to apply nucleation theory to the formation of a solid crystal. In deriving the preceding expressions, we assumed that the shape of an

embryo is not a function of  $n$ , the number of atoms which it contains. Except for very small  $n$  (where the approximation is unimportant), this assumption is justified for liquid droplets, but in a solid crystal, each atom has to fit into a fixed position relative to its neighbours. The configuration presented by the crystal embryo to the vapour thus varies periodically as successive rows and planes of atoms are completed.

Becker and Döring's original paper included a treatment of this problem, assuming the crystal nucleus to be cubical in shape. The expression they derived contained terms for the nucleation of rows and planes of atoms, but these have very little influence on the three-dimensional nucleation rate. We shall show in the next chapter that these terms, moreover, vanish for real crystals. The two-dimensional nucleation of new atomic planes is an important factor in the theory of growth of an *ideal* crystal.

When the extra terms are omitted, Becker and Döring's expression reduces to

$$I = O_1 q_c \frac{\Delta G_c}{kT} \exp\left\{\frac{-\Delta G_c}{kT}\right\}, \quad (49.1)$$

which differs slightly from (48.8) in the pre-exponential term. The change in Zeldovich factor is unimportant, and eqn. (48.14) gives the nucleation rate to sufficient accuracy for most purposes. According to the latest estimate of the replacement partition factor for small crystals (Nishioka *et al.* 1971), the Lothe-Pound correction factor for vapour-crystal nucleation is  $\sim 10^{15}$ , but depends rather strongly on the difference in surface free energy of macroscopic and small crystals.

We must now consider the problem of nucleation in liquid-solid or in solid-solid transformations, i.e. in completely condensed assemblies. The general form of the preceding equations will obviously be maintained, and the nucleation rate will again contain an exponential term involving the increase in free energy for a nucleus of critical size. The pre-exponential term will now be proportional to an encounter rate of atoms in the condensed assembly, but the kinetic theory of gaseous collisions is no longer applicable. As shown in Chapter 9, the encounter rate will be determined essentially by the product of an atomic vibrational frequency and an exponential term containing the activation energy for an elementary atomic movement. This kind of reasoning led Becker (1940) to suggest that the nucleation rate should be written

$$I = C_9 \exp(-\varepsilon/kT) \exp(-\Delta G_c/kT), \quad (49.2)$$

where  $\varepsilon$  is the activation energy for diffusion. The magnitude of the constant  $C_9$  remains unspecified in the general case in Becker's theory, but an explicit value was obtained by Turnbull and Fisher (1949), who applied the reaction rate theory of Chapter 4 to the case where the product phase has the same composition as the parent phase. The simplest example of such a transformation in a condensed assembly is the freezing of a pure metal, and this is considered in the following development of the theory. The theory applies equally to solid-solid transformations, but has to be modified to allow for strain energies.

The excess free energy of an embryo solid crystal is now

$$\Delta G = n(g^s - g^l) + \eta\sigma n^{2/3}, \tag{49.3}$$

where  $\eta$  is a shape factor, as in Section 46. The maximum value of the free energy increase and the size of the critical nucleus are thus

$$\Delta G_c = \frac{4\eta^3\sigma^3}{27(g^l - g^s)^2}, \tag{49.4}$$

$$n_c = \left\{ \frac{2\eta\sigma}{3(g^l - g^s)} \right\}^3. \tag{49.5}$$

These equations are the same as those used previously with  $(g^v - g^l)$  replaced by  $(g^l - g^s)$  and the additional assumption that (46.1) is valid. When both phases are condensed, it seems intuitively improbable that the factors of eqn. (47.6) or (47.7) can be large, so that additional terms in  $\Delta G$ , analogous to those in (46.4), should not be important. However, one reservation must be made in the case of liquid→solid nucleation; Lothe and Pound (1962) have suggested that a  $10^7$  discrepancy between the experimental and theoretical pre-exponential factors might be attributable to nearly free rotation of the solid nuclei within the liquid. This does not seem entirely logical in view of their estimate of  $Q_{rep}^*$ , which depends on the assumption that nearly free rotation of liquid droplets within liquid does not take place.

The growth of an individual embryo may be assumed to involve a large number of small fluctuations, represented in eqns. (47.8), and only rarely will there be a sufficiently long chain of forward fluctuations for an embryo to reach critical size. In the kinetic equation (48.1), for the interchange of  $E_n$  and  $E_{n+1}$  embryos, we now have to find new expressions for the probabilities  $O_n q_0$  and  $O_{n+1} q_{n+1}$  that in unit time embryos  $E_n, E_{n+1}$  will change into each other. We make the reasonable assumption that during the addition

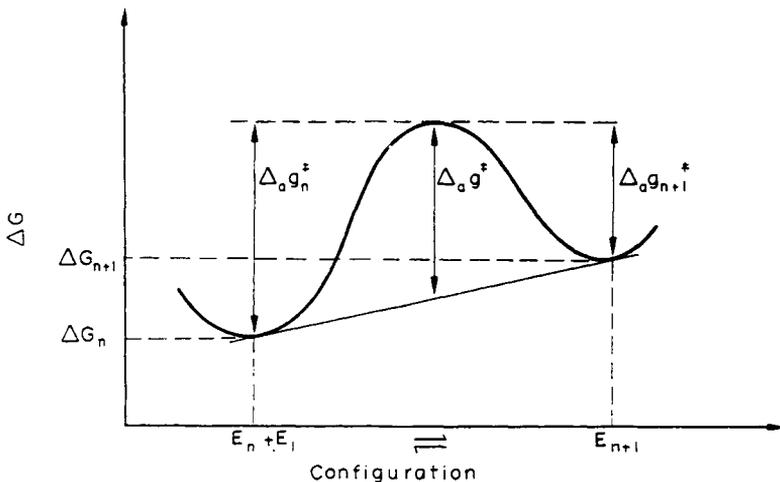


FIG. 10.5. To illustrate energy relations in the Turnbull–Fisher treatment of solid-state nucleation theory.

of an atom to an  $E_n$  embryo, the assembly passes through energy states which are higher than either the initial or final states of the process. There will thus be a maximum free energy at some intermediate stage, and the theory of Chapter 3 can be applied. The energy relations are shown in Fig. 10.5 as a function of the change in configuration from  $E_n$  to  $E_{n+1}$ . The energy  $\Delta G_{n+1}$  is greater than  $\Delta G_n$  since the figure is drawn for  $n < n_c$ . If  $\Delta_a g_n^*$  is the height of the free energy maximum (the activated complex) above the energy  $G + \Delta G_n$  of the assembly containing the  $E_n$  embryo, we may write the rate at which an atom in contact with the embryo will transfer into it as

$$(kT/h) \exp(-\Delta_a g_n^*/kT).$$

The energy has been written as  $\Delta_a g_n^*$  rather than as  $\Delta g_n^*$  (cf. eqn. (13.9)) in order to emphasize the straightforward physical interpretation of eqn. (13.10). As already emphasized, it is preferable to replace  $kT/h$  by a simple frequency  $\nu$ , the magnitude of which is not specified exactly. We retain  $kT/h$  in the equations for nucleation rate in this section, however, to facilitate direct comparison with the equations given by Turnbull and Fisher.

Let the number of atoms in the surface of  $E_n$  be written  $o_n$ . The basic kinetic equation, in place of (48.1), is now

$$I_{n,t} = (kT/h) [Z_{n,t} o_n \exp(-\Delta_a g_n^*/kT) - Z_{n+1,t} o_{n+1} \exp(-\Delta_a g_{n+1}^*/kT)]. \quad (49.6)$$

Writing down the corresponding equation for the equilibrium distribution  $N_n$ , when the liquid phase is stable, we may again express (49.6) in the form

$$I_{n,t} = \frac{kT}{h} \exp\left(\frac{-\Delta_a g_n^*}{kT}\right) N_n o_n \left[ \frac{Z_{n,t}}{N_n} - \frac{Z_{n+1,t}}{N_{n+1}} \right]. \quad (49.7)$$

It is to be noted that  $\Delta_a g_n^*$  is a function of  $n$  only because of the difference in the energies  $\Delta G_n$  and  $\Delta G_{n+1}$ . In fact we may write approximately:

$$\begin{aligned} \Delta_a g_n^* &= \Delta_a g^* + \frac{1}{2}(\partial \Delta G_n / \partial n), \\ \Delta_a g_{n+1}^* &= \Delta_a g^* - \frac{1}{2}(\partial \Delta G_n / \partial n), \end{aligned}$$

where  $\Delta_a g^*$  is the free energy of activation for the transfer of atoms across the interface, and is independent of  $n$ .

The theory is now developed exactly as in Section 48. In the quasi-steady state, we have the same conditions as previously, and by summing the set of equations we obtain

$$\frac{Ih}{kT} \exp \frac{\Delta_a g^*}{kT} \sum_p^s \frac{1}{N_p o_p \exp\{(\frac{1}{2}kT)(\partial \Delta G_n / \partial n)\}} = 1. \quad (49.8)$$

Once again, the only important terms in the sum are those having  $n$  values near  $n = n_c$ . Then  $o_n$  can be taken out of the sum as  $o_c$ , and  $N_n$  is expressed by expanding  $\Delta G_n$  about  $\Delta G_c$ . In all the important terms,  $\partial \Delta G_n / \partial n$  will be very small (it is zero at  $n = n_c$ ), and we may thus equate the exponential term in this quantity to unity in the sum. The summation is thus reduced to (48.7) again, and on substituting into (49.8), the nucleation rate is obtained

as

$$I = N \frac{kT}{h} \frac{o_c}{n_c} \left( \frac{\Delta G_c}{3\pi kT} \right)^{1/2} \exp - \left( \frac{\Delta G_c + \Delta_a g^*}{kT} \right). \quad (49.9)$$

We have obtained (39.9) in a slightly different and simpler manner from that used by Turnbull and Fisher, mainly in order to emphasize that their calculation is essentially a form of the Becker-Döring theory of the last section. According to Turnbull and Fisher, the quantity  $(o_c/n_c) (\Delta G_c/3kT)^{1/2}$  is within one or two powers of ten for all nucleation problems of interest, so eqn. (49.9) may be written with sufficient accuracy as

$$I = N \frac{kT}{h} \exp \left\{ \frac{-(\Delta G_c + \Delta_a g^*)}{kT} \right\}. \quad (49.10)$$

This has the same form as Becker's eqn. (49.2), except that the activation energy involved is that for transfer across an interface, rather than for diffusion. The activation energy for diffusion, however, will clearly be involved if the transformation produces a change of composition; this is dealt with in a later chapter.

The magnitude of the predicted nucleation rate is readily estimated. For the nucleation rate per unit volume, we have

$${}^v I = {}^v N (kT/h) \exp \{ -(\Delta G_c + \Delta_a g^*)/kT \}, \quad (49.11)$$

and reasonable values at ordinary temperatures are  ${}^v N \approx 10^{28} - 10^{29} \text{ m}^{-3}$  and  $kT/h \approx 10^{13} \text{ s}^{-1}$ . The value of  $\Delta_a g^*$  is less certain, but it is estimated by Turnbull that it is approximately equal to the activation energy for viscous flow, giving  $\exp(-\Delta_a g^*/kT) \approx 10^{-2}$  for liquid-solid changes in metals. Thus

$${}^v I \approx 10^{39} \exp(-\Delta G_c/kT) \text{ m}^{-3} \text{ s}^{-1}. \quad (49.12)$$

Some of the published papers give the uncertainty in the numerical factor of this equation as one power of ten, but it is probably two to four powers of ten when allowance is made both for the neglected factors of (49.9) and for the unknown value of  $\Delta_a g^*$ . Since the variations of the exponential term with change in  $\Delta G_c$  is so rapid, the value of  $\Delta G_c$  required to give a fixed nucleation rate is insensitive to the exact value of the pre-exponential term. A reasonable lower limit to the nucleation rate observable under normal conditions (within one or two powers of ten) is  $10^6 \text{ m}^{-3} \text{ s}^{-1}$ , i.e. one nucleus per  $\text{cm}^3$  per s, which corresponds to a free energy of activation  $\Delta G_c \approx 74kT \approx 10^{-18} \text{ J}$  at 1000 K. With  $v^s = 5 \times 10^{-29} \text{ m}^3$  and assuming  $\sigma = 0.1 \text{ J m}^{-2}$  (100 ergs  $\text{cm}^{-2}$ ), (46.19) gives  $g^l - g^s \approx 6 \times 10^{-21} \text{ J/atom}$ , which is equivalent to 150 cal  $\text{cm}^{-3}$  or 1-3 kg cal/mole for most metals. This driving force varies as  $\sigma^{3/2}$ , but should not differ from this estimate by more than one order of magnitude, since  $\sigma$  for most metals probably lies in the range 0.02-0.25 J  $\text{m}^{-2}$  (20-250 ergs  $\text{cm}^{-2}$ ).

In a condensed assembly, the degree of instability of a metastable phase with respect to a stable phase of the same composition is conveniently expressed by the supercooling ( $\Delta T^-$ ) or the superheating ( $\Delta T^+$ ) below or above the thermodynamic transformation temperature. For a liquid-solid change, we have  $g^l = g^s$  at the freezing point ( $T^{ls}$ ), and  $\Delta s^{sl} = \Delta h^{sl}/T^{ls}$ ,

where  $\Delta s^{sl}$  and  $\Delta h^{sl}$  are respectively the entropy and heat of fusion per atom. At a temperature  $T < T^{ls}$ ,

$$g^l - g^s = (\Delta h^{sl})_T - T(\Delta s^{sl})_T,$$

where the heat and entropy of fusion now strictly refer to liquid formed from solid at temperature  $T$ . The heat of fusion, however, only varies with temperature through changes in the  $pV$  terms, which are negligible, and it is also quite a good approximation to regard the entropy of fusion as temperature independent. This gives

$$g^l - g^s = (T^{ls} - T) \Delta s^{sl} = \Delta h^{sl} \frac{\Delta T^-}{T^{ls}}, \quad (49.13)$$

where the supercooling  $\Delta T^- = T^{ls} - T$ .

Substituting into (49.4) and (49.12), the nucleation rate becomes

$${}^vI \simeq 10^{39} \exp \left\{ \frac{-4\eta^3\sigma^3}{27k(\Delta h^{sl})^2T} \frac{(T^{ls})^2}{(\Delta T^-)^2} \right\} \text{m}^{-3} \text{s}^{-1}. \quad (49.14)$$

We cannot easily obtain a general estimate of the value of  $\Delta T^-/T^{ls}$  at which the nucleation rate becomes appreciable, since this is so dependent on  $\sigma$ . It will be seen, however, that if  $\Delta s^{sl}$  has the same value for all metals and if  $\sigma$  is proportional to  $\Delta h^{sl}$ , a given value of  ${}^vI$  would correspond to a constant ratio of  $\Delta T^-/T^{ls}$ . Both these conditions are approximately true, at least for metals of simple crystal structure, and in Part II, Chapter 14, we shall find that experimental results show that the amounts of supercooling which can be obtained with different metals are proportional to the melting points of the metals on the absolute scale. The very rapid variation of nucleation rate with supercooling is illustrated by the fact that changing  $g^l - g^s$  to one-half the value required to give  ${}^vI \simeq 10^6 \text{m}^{-3} \text{s}^{-1}$  reduces the nucleation rate to  ${}^vI \simeq 10^{-90} \text{m}^{-3} \text{s}^{-1}$ .

## 50. TIME-DEPENDENT NUCLEATION

The above solutions to eqns. (48.1) and (49.6) apply only to the limiting conditions in which a quasi-steady distribution of embryos has been established. In many experiments, the assembly is suddenly changed from a stable to a metastable condition, and the nucleation rate is then a function of time until the quasi-steady state is attained. This effect will be unimportant if the transient is of short duration compared with the period of observation, and the steady state solutions (48.8) and (49.10) then provide a good description of the nucleation process. In condensed phases, however, the existence of an activation energy barrier to the addition and removal of atoms from embryos may mean that the quasi-steady distribution  $Z_n$  is only approached slowly, and it is then necessary to investigate the time-dependent nucleation rate. The transient has also been shown to be important in considering vapour condensation effects in supersonic wind tunnels, where the supersaturation is increased extremely rapidly.

Consideration of the embryo distribution and nucleation rate as a function of time requires some specification of the initial state of the assembly at time  $t = 0$ . If the assembly

has been quenched from a stable condition, the initial concentration of embryos of any size will be very small (curve *A* in Fig. 10.6), and a sufficient approximation to this distribution is to assume that there are no embryos present at  $t = 0$ . We have then effectively to calculate how  $Z_{n,t}$  changes from 0 to the quasi-steady  $Z_n$  (curve *C*) as  $t$  increases from 0 to  $\infty$ .

The establishment of the quasi-steady state is represented qualitatively in Fig. 10.6, which shows the number of embryos  $E_n$  as a function of time. With the initial condition of curve *B* in Fig. 10.6, the rate of production of embryos  $E_2$  will obviously be a maximum

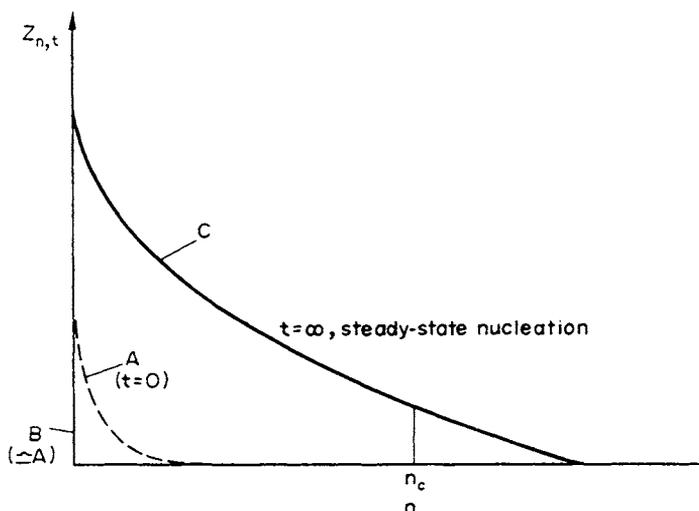


FIG. 10.6. Schematic curves to illustrate the change in embryo distribution with time in a quenched specimen. The initial distribution (*A*) at  $t = 0$  may often be approximated by  $Z_{n,t} = 0$  (*B*).

at  $t = 0$ . For all other embryos ( $n > 2$ ), the rate of production of  $E_n$  will be zero at  $t = 0$ , rising to a maximum at some later time, and then decreasing to zero again as  $Z_{n,t}$  approaches  $Z_n$ . These changes are shown in Fig. 10.7.

Recalling that eqns. (48.1) or (49.6) give the net rate at which embryos  $E_n$  are changed into  $E_{n+1}$ , we see that the rate of change of the number of  $E_n$  embryos present is given by

$$\frac{\partial Z_{n,t}}{\partial t} = I_{n-1,t} - I_{n,t}. \tag{50.1}$$

This set of equations has to be solved subject to the initial conditions

$$\left. \begin{aligned} Z_{1,0} = N, \quad Z_{n,0} = 0 \quad (n \geq 2) \\ \text{and the boundary conditions} \\ Z_{1,t} = N, \quad Z_{s,t} = 0, \quad \text{for all } t. \end{aligned} \right\} \tag{50.2}$$

These boundary conditions are, of course, the same as those used in the steady-state solutions above; the number of atoms in the parent phase is treated as constant (neglecting

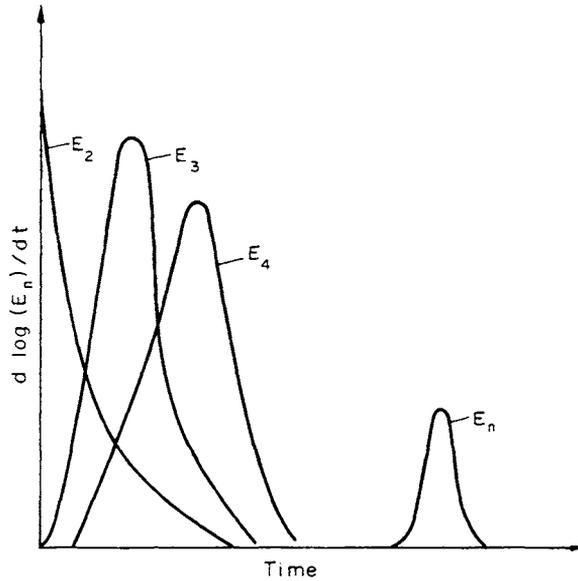


FIG. 10.7. The rate of accumulation of embryos as a function of time (after Turnbull, 1948).

the small decrease shown in Fig. 10.6), and embryos of size  $s$  are assumed to grow rapidly and be effectively removed from the assembly. Even with these assumptions, we were only able to obtain an approximate expression for steady state nucleation, and the problem of finding a complete solution is much more complex. Turnbull (1948) has solved the equations numerically, using suitable values for the parameters. Since the nucleation rate depends essentially on  $Z_c$ , we expect from the above discussion that there will be an induction period

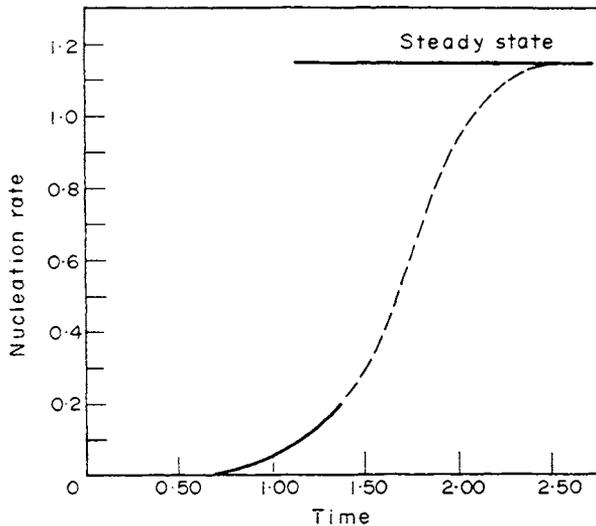


FIG. 10.8. Nucleation rate in an arbitrary example with a critical nucleus size of 25 atoms (after Turnbull, 1948).

during which  $I_t = 0$ , followed by a gradual rise to the steady state value of  $I$ . Turnbull's results, shown in Fig. 10.8, confirm the expected shape of the curve.

Approximate analytical expressions can be obtained only if the difference equations are reduced to differential equations; (48.2) and (49.7) may both be written in the form

$$I_{n,t} = N_n D_n \left[ \frac{Z_{n,t}}{N_n} - \frac{Z_{n+1,t}}{N_{n+1}} \right], \quad (50.3)$$

where  $D_n$  is given the appropriate value, corresponding to nucleation in a vapour or in a condensed phase. The quantities in this equation are defined only at a set of integral values of  $n$ , but we now regard them as functions of a continuous variable  $n$ , this being a good approximation except for very small  $n$ . The equation can thus be written in differential form as

$$I_{n,t} = -N_n D_n \frac{\partial}{\partial n} \left( \frac{Z_{n,t}}{N_n} \right).$$

Substituting from (47.1), this becomes

$$I_{n,t} = -D_n \frac{\partial Z_{n,t}}{\partial n} - \frac{D_n Z_{n,t}}{kT} \frac{\partial \Delta G_n}{\partial n}. \quad (50.4)$$

Finally, from (50.1) and (50.4),

$$\frac{\partial Z}{\partial t} = \frac{\partial}{\partial n} \left( D \frac{\partial Z}{\partial n} \right) + \frac{1}{kT} \frac{\partial}{\partial n} \left( DZ \frac{\partial \Delta G}{\partial n} \right), \quad (50.5)$$

in which we have omitted the subscripts  $n$  and  $t$ .

As pointed out by Zeldovich (1943) and Frenkel (1946), eqn. (50.5) is formally equivalent to the diffusion equation for a set of particles distributed along an axis  $n$  and moving in the influence of a force field specified by a potential  $\Delta G_n$ . It is sometimes called the Fokker-Planck equation, and has been studied in connection with the theory of Brownian motion.  $D_n$  is the formal analogue of a diffusion coefficient, but it is not independent of  $n$ .

The initial and boundary conditions to be satisfied by the solution to (50.5), replacing those of the more exact (50.2), are

$$Z_{n,0} = 0 \quad (n > 0), \quad Z_{0,t} = N, \quad Z_{s,t} = 0. \quad (50.6)$$

The earliest treatment of the problem is due to Zeldovich (1943). He noted that in the region  $n < n_c$ , the exact solution of eqn. (50.5) for the case of harmonically bound particles executing Brownian motion leads to the approximate result

$$Z_{n,t} = Z_n \exp(-n_c^2/4D_c t), \quad (50.7)$$

where  $Z_n$  is the quasi-equilibrium value. Since the nucleation rate will be approximately proportional to  $Z_c$ , this suggests the nucleation current may be written

$$I_t \approx I \exp(-\tau/t) \quad (50.8)$$

and Zeldovich's result is sometimes quoted in this form. It is not usually emphasized, however, that this equation is based on a solution which assumes an invalid relation between  $\Delta G_n$  and  $n$ , and some further justification of (50.8) is thus needed.

Equation (50.5) has also been investigated by Kantrowitz (1951), who pointed out that when  $t$  is small, the second term on the right, representing the thermodynamic work barrier to the formation of embryos, will be relatively unimportant compared with the first term, which represents the kinetic obstacles to growth. An approximate solution can thus be obtained by neglecting this second term, and also treating  $D_n$  as a constant, since its variation with  $n$  is much smaller than that of  $Z_{n,t}$ . Equation (50.5) then reduces to the ordinary diffusion equation

$$\frac{\partial Z}{\partial t} = D \frac{\partial^2 Z}{\partial n^2}$$

and with the appropriate conditions (50.6), the solution is (Carslaw and Jaeger, 1947)

$$Z = N(1 - n/s) - \frac{2N}{\pi} \sum_{m=1}^{\infty} \frac{1}{m} \sin\left(\frac{m\pi n}{s}\right) \exp\left(-\frac{m^2\pi^2 Dt}{s^2}\right). \quad (50.9)$$

The nucleation rate is  $I_{s,t}$ , and from (50.4) this is approximately  $-D(\partial Z/\partial n)_s$ . This gives

$$I_{s,t} = \frac{ND}{2} \left[ 1 + 2 \sum_{m=1}^{\infty} (-1)^m \exp(-m^2\pi^2 Dt/s^2) \right].$$

The series converges slowly at first, but may be transformed by means of the Poisson summation formula (Courant and Hilbert, 1953), into the more rapidly converging form

$$I_{s,t} = 2N(D/\pi t)^{1/2} \sum_{m=1}^{\infty} \exp(-s^2/4Dt(2m-1)^2).$$

For small  $t$ , only the first term need be retained, and

$$I_{s,t} = 2N(D/\pi t)^{1/2} \exp(-s^2/4Dt), \quad (50.10)$$

which is Kantrowitz's expression for the nucleation rate. It shows that the nucleation rate does not become appreciable until a time  $\tau \approx s^2/4D_c \approx n_c^2/D_c$  has elapsed. The result is only valid for small  $t$ , and it is not possible to obtain the relation between  $I_{s,t}$  and the quasi-steady-state nucleation rate  $I$  by allowing  $t$  to tend to infinity in eqn. (50.10). Nevertheless, the general similarity of the expression to Zeldovich's result suggests again that (50.8) may be a reasonable approximation to the nucleation rate in the early part of the transient. It is evident that the duration of the transient is determined by a characteristic time, given to an order of magnitude by

$$\tau = n_c^2/D_c = n_c^2/q_0 O_c. \quad (50.11)$$

A physical argument which leads to a similar approximate value for  $\tau$  has been developed by Russell (1968, 1969). In the range of sizes  $n_1 < n < n_2$  within which  $\Delta G_n + kT \gg \Delta G_c$ , an embryo executes a nearly random walk, i.e. it gains and loses atoms at the same average

rate (Feder *et al.*, 1966). Nuclei of sizes  $n > n_2$  and embryos of sizes  $n < n_1$ , on the other hand, have very small probabilities of shrinkage or growth respectively. Consider a nucleus of size  $n_2$  which ultimately disappears. The time taken will be the sum of the time  $\tau_1$  for the random walk plus the time  $\tau_2$  for the "drift flow", and it can be shown that  $\tau_1 > \tau_2$ . The principle of time reversal requires that these same average times would be those for a nucleus of size  $n_2$  to form *ab initio*, and Russell therefore takes the larger time  $\tau_1$  as the relaxation time  $\tau$  for the establishment of the steady state nucleation rate.

The size range  $n_1 - n_2$  may be derived from the expansion (48.5) which gives

$$n_2 - n_1 = (8kT)^{1/2} (-\partial^2 \Delta G_n / \partial \xi^2)_{\xi=0}^{-1/2}, \tag{50.12}$$

and from (46.3) and (46.20), this becomes†

$$n_2 - n_1 = (-12kT / \Delta G_c)^{1/2} n_c. \tag{50.13}$$

The relaxation time for the random walk is

$$\tau_1 = (n_2 - n_1)^2 / 2q_0 O_c, \tag{50.14}$$

and with the typical values for observable nucleation rates,  $\Delta G_c = 60kT$ , this becomes

$$\tau_1 = n_c^2 / 10q_0 O_c. \tag{50.15}$$

Russell developed this description for a precipitation reaction involving long-range diffusion, in which case the factor  $q_0$  involves the activation energy for volume diffusion, as already emphasized. An alternative estimate by Hillig (1962) takes better account of the long-range diffusion aspects of the problem, but neglects the effects of  $\Delta G_n$ ; it therefore applies only to precipitation from dilute solutions which is considered in a later chapter. Equations (50.11) and (50.15) differ by only one order of magnitude, and greater accuracy cannot be expected; eqn. (50.15) is certainly an underestimate of the relaxation time.

A more complete solution to eqn. (50.5), valid in the important region where  $n$  is not very different from  $n_c$ , has been given by Probstein (1951). The equation is expanded, and the terms in  $\partial \Delta G_n / \partial n$  are omitted, since this is close to zero when  $n$  is near to  $n_c$ . Treating  $D$  as constant, the equation is then reduced to

$$\frac{\partial Z}{\partial t} = D \frac{\partial^2 Z}{\partial n^2} + \left( \frac{D}{kT} \frac{\partial^2 \Delta G}{\partial n^2} \right) Z.$$

The solution to this equation with the conditions (50.5) and the additional assumption  $s = 4n_c/3$  has two terms, the first giving the quasi-equilibrium distribution, and the second being a slowly converging series. As  $\partial^2 \Delta G / \partial n^2 \rightarrow 0$  the solution changes into (50.9). Probstein confirmed that the influence of the work function is relatively unimportant in the early part of the transient; in a typical example, the time to reach a nucleation rate of one-tenth of the quasi-steady rate is 50% greater than predicted by Kantrowitz's equation. Such a

† Russell states that  $(\partial^2 \Delta G_n / \partial \xi^2)_{\xi=0} = 2 \Delta G_c / 3n_c^2$  only for spherical nuclei, but the development of Section 46 shows it to be generally valid within the assumptions used there.

discrepancy is insignificant, since these calculations can only claim to give an order of magnitude to the delay time.

In an isothermal transformation, the significant factor will often be the ratio of the time taken to establish the steady-state nucleation rate to the effective time taken to complete the transformation. When this ratio is small, the transformation kinetics will be described satisfactorily by the assumption that  $vI$  is constant; if it approaches unity, the time dependence of  $vI$  will influence the overall kinetics. From eqn. (50.11) it follows that the ratio increases with  $(\Delta G_c)^2$ , so that transient effects are important when nucleation is difficult. The ratio may also approach unity quite independently of  $\Delta G_c$  if the free energy of activation for growth of a macroscopic region is appreciably less than  $\Delta_g g^*$ . This means that macroscopic regions grow much more rapidly than do embryos, so that the whole transformation is complete before the embryo distribution reaches a steady state. This is further discussed in Section 68.

## 51. HETEROGENEOUS NUCLEATION AND NUCLEATION ON GRAIN BOUNDARIES

In the preceding sections, the formation of a nucleus has been regarded as a homogeneous process occurring with equal probability in all parts of the assembly. In practice, this is unlikely to happen unless the assembly is extremely pure, and also contains (if in the solid state) very few structural defects. More usually, the presence of impurity particles or strained regions of lattice enable nuclei to be formed with a much smaller free energy of activation than that of the homogeneous nuclei. In this section, we describe the modifications to the theory required when nuclei form on foreign particles present in the assembly, on the walls of the container, or at grain boundaries.

When strain energy effects are not important, the catalysing of a nucleation process must depend on a reduction in the net surface energy needed to form a nucleus. This can happen if the formation of an embryo involves the destruction of part of an existing surface, the free energy of which helps to provide the free energy needed for the new surface. The calculations in this section all depend on this assumption.

We begin by making the rather general assumption that the  $\alpha$  phase is in contact with a solid surface,  $S$ , and we calculate the energy required to produce a  $\beta$  embryo, also in contact with  $S$ . If the surface energy,  $\sigma^{\alpha\beta}$ , of the  $\alpha$ - $\beta$  interface is isotropic, the  $\beta$  embryo will be bounded by spherical surfaces of radius  $r$ , except where it is in contact with  $S$ . The volume of the embryo may be written as  $\eta^\beta r^3$ , and its surface area of contact with the  $\alpha$  phase as  $\eta^{\alpha\beta} r^2$ , where  $\eta^\beta$ ,  $\eta^{\alpha\beta}$  are shape factors. The area of contact of the embryo with  $S$  is equal to  $\eta^{\alpha S} r^2$ , the area of  $\alpha$ - $S$  interface destroyed when the  $\beta$  region is formed. The free energy of formation may thus be written

$$\Delta G^S = (\eta^\beta r^3 / v^\beta) (g^\beta - g^\alpha) + r^2 \{ \eta^{\alpha\beta} \sigma^{\alpha\beta} + \eta^{\alpha S} (\sigma^{\beta S} - \sigma^{\alpha S}) \},$$

where  $\sigma^{\alpha\beta}$ ,  $\sigma^{\alpha S}$ ,  $\sigma^{\beta S}$  are the free energies per unit area of the various interfaces. The free energy of formation of the critical size nucleus is found by equating  $\partial \Delta G^S / \partial r = 0$  to find  $r_c$ .

This gives

$$\Delta G_c^S = \frac{4}{27} \frac{\{\eta^{\alpha\beta}\sigma^{\alpha\beta} + \eta^{\alpha S}(\sigma^{\beta S} - \sigma^{\alpha S})\}^3 (\nu^\beta)^2}{(\eta^\beta)^2 (g^\alpha - g^\beta)^2}. \quad (51.1)$$

The problem of heterogeneous nucleation thus reduces mainly to evaluation of  $\eta^\beta$ ,  $\eta^{\alpha\beta}$ ,  $\eta^{\alpha S}$  for various particular cases. The problem is determinate, since surface energy conditions of equilibrium have also to be satisfied whenever three surfaces meet along a line.

The simplest possibility is a solid impurity with a flat surface  $S$ , as in Fig. 10.9. The formation of a stable nucleus of  $\beta$  in contact with  $S$  was first considered by Volmer (1929). A  $\beta$  embryo will obviously be the segment of a sphere, since  $\sigma^{\alpha\beta}$  is assumed constant, and

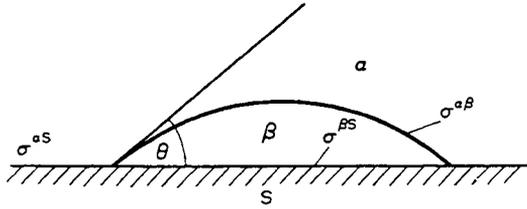


FIG. 10.9. The formation of a  $\beta$  embryo on a flat impurity surface  $S$ .

if the contact angle between the embryo and the surface is  $\theta$ , the condition for static equilibrium is

$$\sigma^{\alpha S} = \sigma^{\beta S} + \sigma^{\alpha\beta} \cos \theta \quad (0 \leq \theta \leq \pi). \quad (51.2)$$

When  $\theta$  lies outside the stated limits, there can be no equilibrium of the surface tension forces, and either the  $\alpha$  or the  $\beta$  phase will spread over the surface.

From the geometry of the figure, we see that  $\eta^\beta = \pi(2 - 3 \cos \theta + \cos^3 \theta)/3$ ,  $\eta^{\alpha\beta} = 2\pi(1 - \cos \theta)$  and  $\eta^{\alpha S} = \pi \sin^2 \theta$ . Hence from (51.1) and (51.2)

$$\Delta G_c^S = \frac{4\pi}{3} \frac{(\sigma^{\alpha\beta})^3 (\nu^\beta)^2 (2 - 3 \cos \theta + \cos^3 \theta)}{(g^\alpha - g^\beta)^2}. \quad (51.3)$$

Comparing with (46.19), we see that the effect of  $S$  is to reduce the formation energy of the corresponding critical spherical nucleus by a factor  $(2 - 3 \cos \theta + \cos^3 \theta)/4$ . For  $\theta > 0$ , this term is always positive, so the presence of the impurity  $S$  cannot enable  $\beta$  embryos to remain stable in the region where  $g^\alpha < g^\beta$ . When  $\theta = \frac{1}{2}\pi$ , the free energy to form a critical nucleus in the interior of the  $\alpha$  phase is twice that to form one on the surface  $S$ . As  $\theta \rightarrow \pi$ , the free energy of the heterogeneous nucleus increases to that required for homogeneous nucleation; as  $\theta \rightarrow 0$ , the free energy decreases to zero. When  $\theta = 0$ , the  $\beta$  phase “wets” the substrate  $S$  in the presence of the  $\alpha$  phase, and the only energy required for the formation of a nucleus is that of its periphery, neglected above. It is believed that this condition is satisfied for the formation of liquid embryos on solid surfaces of the same composition, and that this explains why solids cannot be superheated above their melting points (see Part II, Chapter 14).

If  $\sigma^{\alpha S} > \sigma^{\beta S} + \sigma^{\alpha\beta}$ ,  $\theta$  does not satisfy eqn. (51.2), and there is a negative free energy change in forming the embryo (neglecting the peripheral energy). The embryos in contact

with  $S$  may thus be stable above the thermodynamic transition temperature of the bulk phases,<sup>†</sup> where  $g^\alpha < g^\beta$ . This condition, however, will not often occur.

The nucleation rate for embryos formed on the surface  $S$  may now be found by methods analogous to those used above for homogeneous nucleation. The processes leading to the growth and decay of an embryo are a little uncertain; it seems most probable that the embryos are in quasi-steady equilibrium with atoms from the parent phase which are in contact with  $S$ , or effectively adsorbed on  $S$ . If the number of such atoms is  $N^S$ , then in the region where the  $\alpha$  phase is stable, there will be an equilibrium distribution of embryos of size  $n$

$$N_n = N^S \exp(-\Delta G_n^S/kT).$$

The corresponding steady-state nucleation rate when  $\alpha$  is metastable may thus be written

$$I \simeq N^S q_0 O_r \exp(-\Delta G_c^S/kT), \quad (51.4)$$

for nucleation from a vapour phase, and

$$I \simeq (N^S kT/h) \exp\{-(\Delta G_c^S + \Delta_a g^*)/kT\} \quad (51.5)$$

for nucleation from condensed phases. Thus the only modifications required for changing from homogeneous nucleation equations to those for heterogeneous nucleation consists in substituting  $N^S$  for  $N$  and multiplying  $\Delta G_c$  by a function of a single parameter  $\theta$ . Equation (51.5) can be expressed in the generally more useful form

$${}^S I = (N^S/O^S) (kT/h) \exp\{-(\Delta G_c^S + \Delta_a g^*)/kT\}, \quad (51.6)$$

where  $O^S$  is the total surface area of  $S$  in the assembly, and  ${}^S I$  is the nucleation rate per unit area of  $S$ . The nucleation rate per unit volume is clearly

$${}^v I^S = {}^v O^S {}^S I = {}^v N^S (kT/h) \exp\{-(\Delta G_c^S + \Delta_a g^*)/kT\}, \quad (51.7)$$

where  ${}^v O^S$ ,  ${}^v N^S$  are respectively the surface area and number of surface atoms of  $S$  present in unit volume. Note that whereas the homogeneous nucleation rate is proportional to the volume of the assembly, the heterogeneous rate is proportional to the surface area of the impurity which catalyses the transformation. Thus the state of dispersion of the impurity, as well as the total amount, is effective in determining the nucleation rate.

A uniform flat substrate is not a good approximation to practical conditions. Surface roughening will always be present to an appreciable extent, even on a supposedly smooth flat surface, and the possibility of having  $\beta$  embryos present in surface cavities was first pointed out by Volmer (1939). For  $\theta < \frac{1}{2}\pi$ ,  $\sigma^{\alpha S} > \sigma^{\beta S}$  (eqn. (51.2)), and the part of the surface energy change due to the substitution of the  $\beta$ - $S$  surface for the previous  $\alpha$ - $S$  surface is thus negative. If the embryo can be so shaped that its surface is bounded mainly by  $S$ , rather than by  $\alpha$ , the total surface energy change may become negative. The ratio of the  $\alpha$ - $\beta$  surface to the  $\beta$ - $S$  surface is decreased when the embryo fills a cavity in  $S$ , and under suitable conditions, such an embryo may remain stable at temperatures considerably

<sup>†</sup> We are assuming for definiteness that  $\alpha$  is stable at higher temperatures than  $\beta$ . This and subsequent statements are equally applicable below the transformation temperature if  $\alpha$  is the low-temperature phase.

above the thermodynamic transformation temperature. The retention of embryos in cylindrical and conical cavities has been considered in some detail by Turnbull (1950). He has used his results to explain the variation of the rate of nucleation in liquid–solid transformations with the extent to which the liquid is previously heated above the melting point.

As an illustration of the effect of nucleation in a cavity, consider a cylindrical hole of radius  $r$  containing a height  $h$  of the  $\beta$  phase (Fig. 10.10). We then have  $\eta^\beta = \pi h/r$ ,  $\eta^{\alpha\beta} =$

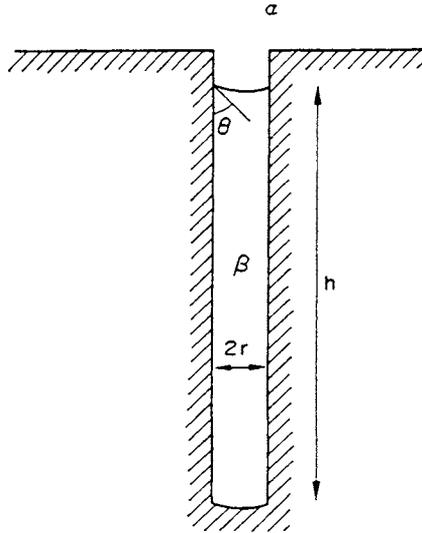


FIG. 10.10. The formation of  $\beta$  phase in a cylindrical cavity.

$2\pi(1 - \sin \theta)/\cos^2 \theta$  and  $\eta^{\alpha\beta} = \pi(2h + r)/r$ . The change in free energy after formation of the embryo is thus

$$\Delta G = \pi r^2 h (g^\beta - g^\alpha) / v^\beta + 2\pi r \sigma^{\alpha\beta} \{ r(1 - \sin \theta) / \cos^2 \theta - (h + \frac{1}{2}r) \cos \theta \}. \quad (51.8)$$

If  $h$  is sufficiently large, the surface energy term in (51.8) may be negative. The stability of an embryo in a cavity above the transformation temperature, however, does not depend on the sign of  $\Delta G$ , but on that of the coefficient of the term in  $h$ .<sup>†</sup> This is because the term in  $r$  will be unchanged when a small quantity of the embryo is absorbed into the  $\alpha$  phase, and the change in free energy thus produced will depend only on  $\Delta h$ . Clearly, the embryo will be stable if the change in  $\Delta G$  due to a decrease in  $h$  is positive, i.e. if

$$\left. \begin{aligned} &(\pi r^2 / v^\beta) (g^\beta - g^\alpha) - 2\pi r \sigma^{\alpha\beta} \cos \theta < 0 \\ \text{or} \quad &r < 2v^\beta \sigma^{\alpha\beta} \cos \theta / (g^\beta - g^\alpha). \end{aligned} \right\} \quad (51.9)$$

The smaller the cavity radius  $r$ , the greater is the chance of the embryo remaining stable.

<sup>†</sup> This statement is, of course, true only in a limited sense. In principle, an embryo will ultimately disappear if  $\Delta G$  is positive, but the probability of this happening in any reasonable time is nearly negligible unless  $\Delta G$  decreases when a small quantity of the embryo disappears.

As the temperature is increased above the transformation temperature,  $(g^\beta - g^\alpha)$  increases, and more and more cavities become unstable, until eventually all are empty.

Once a cavity is empty of  $\beta$ , a  $\beta$  embryo can fill it again only by first forming on the nearly flat bottom or sides of the hole. The critical free energy for this process is thus given not by (51.8) but by (51.1), so that embryos in cavities will not form again until  $(g^\beta - g^\alpha)$  is sufficiently negative for eqns. (51.4) or (51.5) to give an appreciable rate.

The above discussion of heterogeneous nucleation applies in principle to all transformations, but in solid-solid transformations, impurity particles are probably less important than grain boundaries or structural defects as nucleation sites. If strain energy effects are neglected, grain boundary nucleation may be treated as a simple extension of the above problems. The only surface energies involved are now  $\sigma^{\alpha\beta}$  and  $\sigma^{\alpha\alpha}$ , the grain boundary energy of the  $\alpha$  phase.† In eqn. (51.1) the term in  $\sigma^{\beta\beta}$  disappears, and  $\eta^{\alpha\beta}$ ,  $\sigma^{\alpha\beta}$  are replaced by  $\eta^{\alpha\alpha}$ ,  $\sigma^{\alpha\alpha}$  respectively.

We have to consider the possibility of nuclei forming at surfaces separating two grains, at edges where three grains meet, or at corners common to four grains. For a two-grain boundary, the embryo (Fig. 10.11) will be a symmetrical doubly-spherical lens, so that

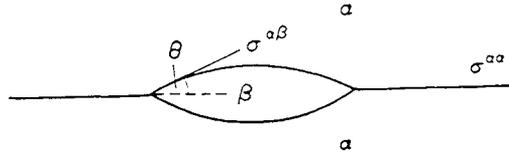


FIG. 10.11. The formation of a  $\beta$  embryo on an  $\alpha$  grain boundary surface (after Clemm and Fisher, 1955).

$\eta^\beta = 2\pi(2 - 3 \cos \theta + \cos^3 \theta)/3$ ,  $\eta^{\alpha\beta} = 4\pi(1 - \cos \theta)$  and  $\eta^{\alpha\alpha} = \pi \sin^2 \theta$ , where  $\theta$  is the contact angle as before. The condition for static equilibrium is

$$\sigma^{\alpha\alpha} = 2\sigma^{\alpha\beta} \cos \theta. \quad (51.10)$$

The free energy of formation of the critical nucleus is thus found to be twice that given by eqn. (51.3), and the ratio of this energy to the energy for homogeneous nucleation is

$$\Delta G_c^B / \Delta G_c^H = \frac{1}{2}(2 - 3 \cos \theta + \cos^3 \theta), \quad (51.11)$$

where we have written  $\Delta G_c^B$  for the critical free energy needed to form a nucleus on the boundary, and  $\Delta G_c^H$  for the critical free energy for homogeneous nucleation.

At a three-grain junction, we assume the equilibrium configuration (p. 331) in which three planar boundaries meet in a line at angles of  $120^\circ$  to each other. The shape of an embryo will then be a figure bounded by three spherical surfaces; a section normal to the line is shown in Fig. 10.12. From this we see that the condition for static equilibrium is

† The treatment in this section assumes that  $\sigma^{\alpha\beta}$  is isotropic. Recent work on grain boundary nucleation (Lee and Aaronson, 1975; Johnson *et al.*, 1975; Lange and Aaronson, 1979) emphasises the possible importance of nuclei with lattices so orientated that particular planar interfaces have very low energies. The nuclei then adopt pillbox or other shapes in which most of the interface consists of such low energy planes, and  $\Delta G_c^B$  is correspondingly reduced.

again given by eqn. (51.10), where  $\theta$  is the dihedral angle at the edge between two  $\alpha$ - $\beta$  surfaces and an  $\alpha$ - $\alpha$  surface, as before. The geometry is a little more tedious, but it is readily seen that if the radius of the spherical surfaces is  $r$ , the length of the edge along which the embryo forms is  $2r(1 - 4 \cos^2 \theta/3)^{1/2}$ . The three edges, in each of which two spherical

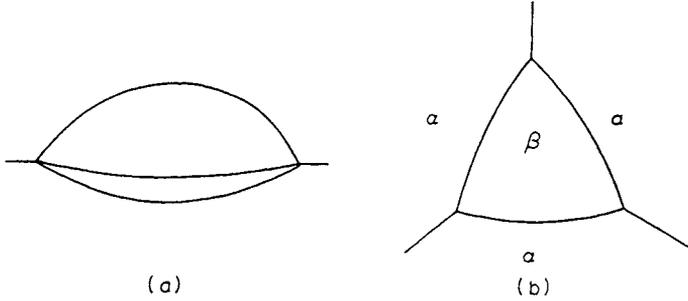


FIG. 10.12. The shape of a  $\beta$  embryo formed at a three-grain junction in  $\alpha$ . (a) General view, (b) section normal to grain edge (after Clemm and Fisher, 1955).

boundaries of the embryo meet a planar grain boundary, are small circles of radius  $r \sin \theta$ . Calculation of the volume and of the planar and spherical surface areas then gives

$$\left. \begin{aligned} \eta^\beta &= 2\left[\pi - 2 \arcsin\left(\frac{1}{2}\operatorname{cosec} \theta\right) + \frac{1}{3}\cos^2 \theta(4 \sin^2 \theta - 1)^{1/2} \right. \\ &\quad \left. - \arccos(\cot \theta/\sqrt{3}) \cos \theta(3 - \cos^2 \theta)\right], \\ \eta^{\alpha\beta} &= 6\pi - 12 \arcsin\left(\frac{1}{2}\operatorname{cosec} \theta\right) - 12 \cos \theta \arccos(\cot \theta/\sqrt{3}), \\ \eta^{\alpha\alpha} &= 3 \sin^2 \theta \arccos(\cot \theta/\sqrt{3}) - \cos \theta(4 \sin^2 \theta - 1)^{1/2}. \end{aligned} \right\} \quad (51.12)$$

From this,  $\eta^{\alpha\beta} - 2\eta^{\alpha\alpha} \cos \theta = 3\eta^\beta$ , and hence

$$\Delta G_c^E = 4\eta^\beta \sigma^{\alpha\beta} (v^\beta)^2 / (g^\beta - g^\alpha)^2.$$

The ratio of the nucleation energy to that for homogeneous nucleation is now

$$\Delta G_c^E / \Delta G_c^H = (3/4\pi)\eta^\beta, \quad (51.13)$$

where  $\eta^\beta$  is given by (51.12) above. We should note that in deriving  $\Delta G_c^E$  we have neglected any extra edge energy, and have assumed that the energy gain is given entirely by the free energies of the planar boundaries eliminated.

Finally, we have to consider a grain corner where four different grains meet. The four grain edges, each being the junction of three of the grains, may be assumed to radiate symmetrically from the corner, so that an embryo bounded by spherical surfaces will have the shape of a spherical tetrahedron (Fig. 10.13).<sup>†</sup> When the volume and area shape factors are

<sup>†</sup> It will be realized that this condition, and the condition assumed for edge nucleation, are not likely to be exactly realized in practice, for reasons discussed on pp. 332-4. However, the assumption is a good approximation to the average actual configuration.

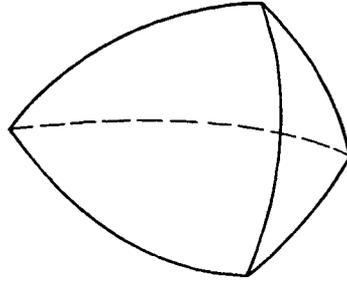


FIG. 10.13. Possible shape of a  $\beta$  embryo forming at a grain corner (after Clemm and Fisher, 1955).

evaluated, it is found that once again  $\eta^{\alpha\beta} - 2 \cos \theta \eta^{\alpha\alpha} = 3\eta^\beta$ , so that  $\Delta G_c^C$  is also given by (51.13), with the appropriate value of  $\eta^\beta$ , which is

$$\eta^\beta = \left. \begin{aligned} & [8\{\pi/3 - \arccos\{[\sqrt{(2) - \cos \theta (3 - C_{10}^2)^{1/2}}]/C_{10} \sin \theta\}\}] \\ & + C_{10} \cos \theta \{(4 \sin^2 \theta - C_{10}^2)^{1/2} - C_{10}^2/\sqrt{2}\} \\ & - 4 \cos \theta (3 - \cos^2 \theta) \arccos C_{10}/(2 \sin \theta) \end{aligned} \right\} \quad (51.14)$$

where  $C_{10} = 2\{\sqrt{(2) (4 \sin^2 \theta - 1)^{1/2} - \cos \theta}\}/3$ .

The ratio of the free energy required to form a grain boundary nucleus to that needed to form a homogeneous nucleus obviously decreases as the ratio of the grain boundary energy to the interphase boundary energy increases. This ratio, given by eqn. (51.13) with appropriate values of  $\eta^\beta$ , is plotted in Fig. 10.14 as a function of  $\cos \theta = \frac{1}{2} \sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ . It will be noted that for all values of  $\cos \theta$ ,  $\Delta G_c^C < \Delta G_c^E < \Delta G_c^B < \Delta G_c^H$ . The ratio of the energies

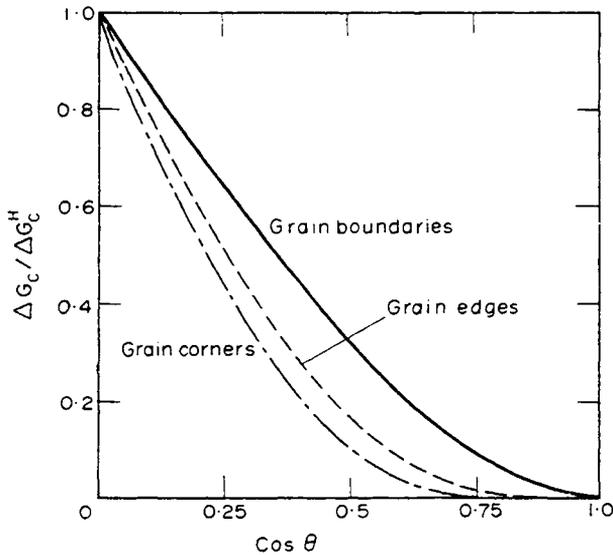


FIG. 10.14. The ratio of the free energy required to form a nucleus on various types of grain boundary site to that required to form a nucleus in the interior of a grain is plotted as a function of  $\cos \theta = \frac{1}{2} \sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$  (after Cahn, 1956).

becomes zero at some finite value of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ , and for boundary energies which are relatively higher than this, no equilibrium is possible. There is then a continuous energy decrease as the new phase grows from zero size, and the amount of transformation depends only on a growth rate. From Figs. 10.11, 10.12, and 10.13 we see that the minimum values of  $\theta$  which allow surface energy equilibrium correspond to infinite radii of the spherical surfaces, and are  $0^\circ$ ,  $30^\circ$ , and  $\sin^{-1}(1/\sqrt{3})$ . The critical values of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ , above which  $\Delta G_c$  is zero, are thus 2,  $\sqrt{3}$  and  $2\sqrt{2}/\sqrt{3}$  for boundaries, edges, and corners respectively.

The relative nucleation rates do not necessarily increase in the same order as the activation energies for nucleation decrease, since the density of sites also decreases as the mode of nucleation changes from homogeneous to grain corners. We have already seen that for boundary nucleation,  $N$  in eqn. (49.10) must be replaced by  $N^B$ , and we may similarly expect factors  $N^E$ ,  $N^C$  for edge and corner nucleation. It must be admitted that these pre-exponential factors are rather ill-defined, since the way in which atoms within an embryo forming at a surface, edge, or corner interchange with other atoms is somewhat uncertain. However, if we suppose that the effective thickness of a grain boundary is  $\delta^B$ , and the mean grain diameter is  $L^B$ , we can write the number of atoms per unit volume on the various sites as

$$\left. \begin{aligned} v_{N^B} &= vN(\delta^B/L^B), \\ v_{N^E} &= vN(\delta^B/L^B)^2, \\ v_{N^C} &= vN(\delta^B/L^B)^3, \end{aligned} \right\} \quad (51.15)$$

to a fair approximation. The nucleation rate per unit volume due to grain boundary surfaces is then related to the corresponding homogeneous nucleation rate  $v_{I^H}$  by

$$\frac{v_{I^B}}{v_{I^H}} = \frac{\delta^B}{L^B} \exp\left\{ \frac{\Delta G_c^H - \Delta G_c^B}{kT} \right\} \quad (51.16)$$

with similar equations for  $v_{I^E}$  and  $v_{I^C}$ . From these equations we find the conditions under which homogeneous nuclei, grain boundary nuclei, grain edge nuclei and grain corner nuclei respectively make the greatest contribution to the overall volume nucleation rate  $v_I$ . These conditions may readily be expressed in terms of a quantity  $R^B = kT \ln(L^B/\delta^B)$  and are:

Greatest nucleation rate	
$v_{I^H}$	$R^B > \Delta G_c^B - \Delta G_c^B$
$v_{I^B}$	$\Delta G_c^H - \Delta G_c^B > R^B > \Delta G_c^B - \Delta G_c^B$
$v_{I^E}$	$\Delta G_c^B - \Delta G_c^E > R^B > \Delta G_c^B - \Delta G_c^C$
$v_{I^C}$	$\Delta G_c^E - \Delta G_c^C > R^B$

The corresponding limits on  $R^B/\Delta G_c^H$  may be expressed in terms of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$  by means of Fig. 10.14. Figure 10.15 shows these regions for different values of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ , as calculated in this way by Cahn.

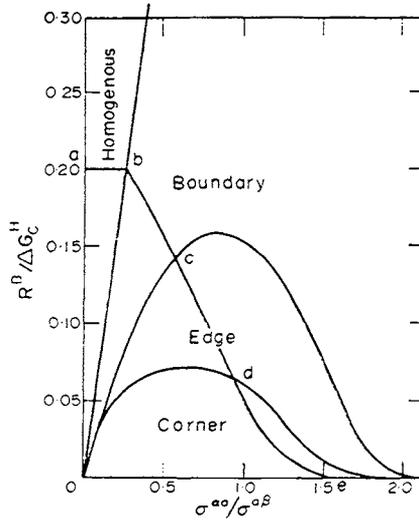


FIG. 10.15. To illustrate the conditions under which various types of site make the greatest contribution to the initial nucleation rate (after Cahn, 1956). The curve *abcde* represents the conditions for the minimum observable nucleation rate. The type of nucleation first observed will depend upon the region in which this curve lies for the given  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ .

It remains to examine the magnitude of the predicted effects. For homogeneous nucleation in the solid, reasonable values of the parameters give eqn. (49.11) as

$$v_I^H \approx 10^{36} \exp(-\Delta G_c^H/kT) \text{ m}^{-3} \text{ s}^{-1} \quad (51.17)$$

and for  $v_I^H \sim 10^6 \text{ m}^{-3} \text{ s}^{-1}$ , we thus obtain  $\Delta G_c^H \approx 69 \text{ kT}$ . If the grain diameter is  $\sim 0.1 \text{ mm}$ ,  $L^B/\delta^B \approx 10^6$ , and the boundary nucleation rate is

$$v_I^B \approx 10^{30} \exp(-\Delta G_c^B/kT) \text{ m}^{-3} \text{ s}^{-1}. \quad (51.18)$$

so that for  $v_I^B \sim 10^6 \text{ m}^{-3} \text{ s}^{-1}$ ,  $\Delta G_c^B \approx 57 \text{ kT}$ . For any particular ratio of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ , the ratio  $\Delta G_c^B/\Delta G_c^H$  of the activation energies for boundary and homogeneous nucleation under the same driving force may be read from the upper curve of Fig. 10.15. From the value of  $\Delta G_c^H$  when  $\Delta G_c^B = 57 \text{ kT}$ , we can plot a curve of  $kT \ln(L^B/\delta^B)/\Delta G_c^H$  (assuming the above parameters) against  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$  on Fig. 10.15. The curve gives the relation between  $\Delta G_c^H$  and  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$  needed to sustain the assumed nucleation rate; part of it is shown as the curve *b-c* on the figure. When the value of  $\Delta G_c^H$  determined in this way equals  $69 \text{ kT}$ , the homogeneous and boundary volume nucleation rates are both equal to the assumed unit nucleation rate, and this gives the point *b* on the figure. At values of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$  smaller than this, a smaller value of  $\Delta G_c^H$  (i.e. a larger driving force  $g^\alpha - g^\beta$ ) is needed to give  $v_I^B \sim 10^6 \text{ m}^{-3} \text{ s}^{-1}$ , but this rate can of course be obtained by homogeneous nucleation at a driving force corresponding to  $\Delta G_c^H = 69 \text{ kT}$ .

For grain-edge and grain-corner nucleation we have the corresponding equations

$$\left. \begin{aligned} vI^E &\approx 10^{24} \exp(-\Delta G_c^E/kT) \text{ m}^{-3} \text{ s}^{-1}, \\ vI^C &\approx 10^{18} \exp(-\Delta G_c^C/kT) \text{ m}^{-3} \text{ s}^{-1}, \end{aligned} \right\} \quad (51.19)$$

giving  $10^6$  nuclei  $\text{m}^{-3} \text{s}^{-1}$  when  $\Delta G_c^H \approx 41kT$  and  $\Delta G_c^C \approx 28kT$  respectively. The corresponding values of  $\Delta G_c^H$  for various ratios of boundary energies may be derived from Fig. 10.14, and the corresponding values of  $kT \ln(L^B/\delta^B)/\Delta G_c^H$  are plotted in Fig. 10.15 over the regions  $cd$  (for edges) and  $de$  (for corners). The whole curve  $abcde$  thus shows the maximum value of  $\Delta G_c^H$ , and hence (using eqn. (46.19)) the minimum value of  $g^\alpha - g^\beta$ , which will allow the just perceptible nucleation rate of  $10^6$  nuclei  $\text{m}^{-3} \text{s}^{-1}$  to be attained. It follows from the curve that the type of nucleation which occurs when  $g^\alpha - g^\beta$  becomes large enough to give a measurable rate of transformation depends on the value of  $\sigma^{\alpha\alpha}/\sigma^{\alpha\beta}$ . In principle, if  $g^\alpha - g^\beta$  is slowly increased from zero (e.g. by continuous cooling), nucleation will initially be greatest on corners, then on edges, then boundaries, and then homogeneously. But if  $\sigma^{\alpha\alpha} < 0.9\sigma^{\alpha\beta}$ , the corner nucleation rate will be too small to be observed until the driving force has increased to such a value that edge nucleation is more rapid than corners; if  $\sigma^{\alpha\alpha} < 0.6\sigma^{\alpha\beta}$ , the amount of edge nucleation will similarly be too small to be observed before boundary nucleation has become predominant; and if  $\sigma^{\alpha\alpha} < 0.25\sigma^{\alpha\beta}$ , only effectively random homogeneous nucleation will ever be observed.†

## 52. NUCLEATION IN THE SOLID STATE

A change of volume accompanies most phase transformations, and its effects have specifically to be considered when the transformation occurs entirely in the solid state. If the parent phase is a gas or a liquid, the volume change is achieved with negligible increase in energy by flow of the surrounding fluid. When a small region in a crystal transforms, however, the rigidity of the material may enable it to withstand very considerable stresses with negligible flow or creep rate. The volume change has then to be accommodated in the assembly, and the associated strain energy is an important factor in the transformation. In this section, we consider its importance in nucleation theory.

We consider the following sequences of imaginary operations, leading to the production of a small  $\beta$  crystal in the middle of a large amount of parent  $\alpha$  material.

- (1) Remove a small volume from the centre of the  $\alpha$  and allow it to undergo an unconstrained transformation to  $\beta$ .
- (2) Apply surface tractions to the  $\beta$  to return it to its original size and shape, and insert it into the hole in the  $\alpha$  again.
- (3) Weld together the  $\alpha$  and (constrained)  $\beta$  regions over their surface of contact.
- (4) Allow the assembly to relax; i.e. remove the built-in layer of surface force by applying an equal and opposite layer of surface force.

The final assembly of matrix  $\alpha$  + small  $\beta$  crystal is clearly in a state of self-stress, the strain energy of which is a factor influencing the nucleation rate. The operations we have specified lead to a  $\beta$  nucleus which is coherent with the  $\alpha$ , in the sense that the displacements and

† Theoretical and experimental work by Aaronson and his collaborators (see, for example Lee and Aaronson (1975)) shows that embryos of the types shown in figures 10.9 and 10.11 cannot account for the observed nucleation rate on grain boundaries. Nucleation on impurity particles and grain boundaries probably utilises special interfaces of low energy and the nucleus shape then contains at least one planar interface of this type.

transformations are continuous across the common interface. The unconstrained transformation to  $\beta$  has been assumed to occur without mass transfer, so that there is a one-to-one correspondence between the atomic positions in the original  $\alpha$  and final  $\beta$  crystals. When the final  $\beta$  crystal approximates to a flat plate, we may usefully distinguish two types of coherence. The first possibility is that the principal axes of the matrix which specifies the free transformation  $\alpha \rightarrow \beta$  lie in the plane of the plate and perpendicular to it. In the final state, much of the strain will remain in the  $\beta$  particle, which will be extended or compressed so as to match the  $\alpha$  phase along their common boundaries. This type of coherent nucleus is believed to be formed in many nucleation and growth reactions. Alternatively, the principal axes may be inclined to the plate in such a way that the displacements correspond to a shear on the plane of the plate, possibly combined with an expansion or contraction normal to the plate. The strain in the final state will then consist largely of shears in the surrounding matrix. This type of coherency is characteristic of martensitic reactions, but may also occur in the nucleation stage of other transformations.

The formation of a small  $\beta$  crystal with an incoherent interface is represented by the above steps if we assume the shear modulus of the transformed region to be zero. The appropriate model is thus to make a hole in the  $\alpha$  and fill it with a compressible fluid, of natural volume equal to the volume of the freely transformed  $\alpha$ . The only condition now being imposed is that the same total number of atoms are removed from and reinserted into the hole; there is no correspondence of atomic positions. Physically, this means that atoms will migrate so as to minimize the energy; in a flat plate of a  $\beta$  phase having larger specific volume than the  $\alpha$  phase, for example, atoms migrate from the edges to the centres of the faces, which can readily bulge outwards. In a sense, the formation of an incoherent interface requires the continuous recrystallization of the  $\beta$ , and will occur at temperatures such that the atoms can migrate fairly rapidly under the (very large) transformation stresses.

If the  $\beta$  particle is a sphere, the distinction between the coherent and incoherent particles disappears. However, we shall find that the strain energy of an incoherent  $\beta$  particle is a function of its shape, and in suitable conditions can be made very small. The strain energy of a coherent particle also varies with its shape, but only a limited reduction can usually be achieved by changing the shape. Thus we might expect that all nuclei will form incoherently except when the volume change is very small. We have noted previously, however, that the surface energy of an incoherent boundary is much larger than that of a coherent boundary, and this opposes the strain energy factor. Since the strain energy is proportional to the volume of the  $\beta$  crystal, the surface energy term will predominate at sufficiently small sizes, and the first nuclei may be coherent. As the nucleus grows, the strain energy will increase until it becomes more favourable energetically for the particle to "break away" from the matrix. This breaking away will usually occur quite early in a nucleation and growth reaction, providing the temperature is high enough to ensure that equilibrium is eventually reached. It cannot happen in a martensitic reaction since the mechanism of continued growth in this case depends on the maintenance of coherence.

All the above virtual processes lead to some strain energy unless the volume change is zero. Zero strain energy is obtained only if we imagine that the original hole in the  $\alpha$  is enlarged or reduced by the removal or addition of slices of  $\alpha$  material until the freely trans-

formed region just fits into it again. There is then no built-in surface force layer, and  $\alpha$  and  $\beta$  are both unstressed, although there may be atomic readjustments at the interface to minimize the surface energy. This is the process used in defining the interfacial energy (Section 38), and the elastic energy discussed in this section is understood to be the energy of the constrained assembly additional to this quantity. The total number of atoms in the region which is finally  $\beta$  is now not conserved, and the possibility of producing a  $\beta$  region without strain energy in this way depends on whether diffusion or plastic flow can readily accommodate the changes in size and shape.

At sufficiently high temperatures, flow rates in a crystal should be sufficient to relieve all but the smallest transformation stresses, and strain energy factors will not be important. At low temperatures, where diffusion rates are slow and ineffective, some strain energy will be created, and the magnitude will be determined by the maximum stress which can be supported by the  $\alpha$  phase before plastic flow begins. This is not an easy problem. The ordinary yield stress criteria are of little value, since a nucleus is so small that there is a negligible chance of its containing a Frank-Read source (this only applies, of course, if it is formed randomly within the assembly; heterogeneous nucleation on preferred sites, including dislocations, is considered later). The surrounding matrix will thus probably be perfect crystal, and should support a shear stress about 1000 times greater than that corresponding to the activation of a dislocation source. If a nucleus does form in a perfect crystal in this way, the transformation stresses will not usually be able to activate Frank-Read sources in the vicinity, since they are applied over such a small volume ( $\sim 10^{-26}$  m<sup>3</sup>).

The conclusion suggested by the above arguments is that the strain energy for nucleation is obtained by assuming the nucleus to be formed in a region of entirely good crystal. However, although single dislocations cannot readily reduce stresses developed in small volumes, certain larger defects such as grain boundaries may be able to do this. In cases where the transformation strain energy in a perfect crystal would be very large, heterogeneous nucleation at places where local flow can occur is to be anticipated, since at these sites the work of nucleation is so much reduced. This effect is additional to the reduction in the surface energy term, discussed in the last section. It shows, however, that when grain boundary nucleation is to be considered, the treatment of the last section is probably adequate, even with considerable transformation volume changes.

We shall now give the standard nucleation theory for a transformation where strain energy effects are not negligible. As in earlier sections, we find it convenient to develop the theory for homogeneous nucleation, since the modifications to allow for nucleation at preferred sites are easily made. The formation of a nucleus of given size will now require an increase in free energy given by (46.3) with the addition of the elastic energy term. The elastic energy is proportional to the number of atoms in the nucleus, so that we may write

$$\Delta G = n(g^\beta - g^\alpha + \Delta g_s) + \eta \sigma n^{2/3}, \quad (52.1)$$

where  $\Delta g_s$  is the elastic energy per atom. In contrast to the previous treatment, the shape of the nucleus is not determined by the condition that the surface energy term shall have its lowest value, since this shape also affects the value of  $\Delta g_s$ . If  $\Delta G$  is plotted as a function of size and shape, we have to find the combination of size and shape which gives a "mini-

max" (saddle-point) value to  $\Delta G$ . The value  $\Delta G_c$  at this point in the energy field represents the activation energy for nucleation by the most favourable path from  $\alpha$  to  $\beta$ .

From eqn. (52.1), we see that the elastic energy reduces the effective driving force ( $g^\alpha - g^\beta$ ) of the reaction. Clearly, nucleation will not occur at all unless

$$g^\alpha - g^\beta > \Delta g_s.$$

Consider first the formation of an incoherent  $\beta$  nucleus. If the  $\beta$  crystal is spherical, the misfitting sphere model of Section 25 may be applied immediately to give the strain energy. The continuum approximation obviously has much more justification in the present application, where the inclusion really is spherical, and has better defined elastic properties; the main limitation is now in the neglect of crystal anisotropy, made in the interests of simplicity. From eqn. (25.16) we see that the strain energy per atom of nucleus is given by

$$\Delta g_s = 2\mu^\alpha C_6 (v^\beta - v^\alpha)^2 / 3v^\beta, \quad (52.2)$$

where  $C_6 = 3K^\beta / (3K^\beta + 4\mu^\alpha)$ , and  $v^\alpha, v^\beta$  are the specific volumes of atoms in the  $\alpha$  and  $\beta$  phases.

When the nucleus has some other shape, we may proceed as follows. Let  $\Delta g_s^\alpha, \Delta g_s^\beta$  be the strain energies per atom when the *whole* of the volume misfit  $3\varepsilon$  is taken up respectively by the surrounding  $\alpha$  matrix or by compression (expansion) of the  $\beta$  particle. Then if the actual volume of the  $\beta$  nucleus is  $(1 + 3C_6\varepsilon)$  times that of the hole, the compression in it is  $3(C_6 - 1)\varepsilon$ . The energy in the particle is  $n(C_6 - 1)^2 \Delta g_s^\beta$ ; and that in the matrix is  $nC_6^2 \Delta g_s^\alpha$ . The total strain energy is thus given by

$$\Delta g_s = (C_6 - 1)^2 \Delta g_s^\beta + C_6^2 \Delta g_s^\alpha$$

and this has its minimum value when

$$C_6 = \Delta g_s^\beta / (\Delta g_s^\alpha + \Delta g_s^\beta)$$

giving

$$\Delta g_s = \Delta g_s^\alpha \Delta g_s^\beta / (\Delta g_s^\alpha + \Delta g_s^\beta) = C_6 \Delta g_s^\alpha = (1 - C_6) \Delta g_s^\beta. \quad (52.3)$$

In particular, if  $\Delta g_s^\beta \gg \Delta g_s^\alpha$ ,  $\Delta g_s \simeq \Delta g_s^\alpha$  and the inclusion is virtually unstrained. We note that  $C_6$  is a measure of the partition of energy between the matrix and the  $\beta$  nucleus, which is

$$\frac{\text{Energy in } \alpha}{\text{Energy in } \beta} = \frac{C_6^2 \Delta g_s^\alpha}{(C_6 - 1)^2 \Delta g_s^\beta} = \frac{C_6}{1 - C_6} \quad (52.4)$$

in agreement with eqns. (25.14) and (25.15) for the case of a sphere. From (52.3),  $C_6$  is also a measure of the extent to which the matrix is able to accommodate the volume change. Since the strain energy is  $(1 - C_6)$  times its value in an incompressible medium, good accommodation is specified by values of  $C_6$  approaching unity, and most of the remaining energy then resides in the matrix.

For a spherical particle, the energy when all the strain is taken by the matrix is

$$\Delta g_s^\alpha = 2\mu^\alpha (v^\beta - v^\alpha)^2 / 3v^\beta.$$

Nabarro proposed that for the more general shapes represented by the family of ellipsoids of revolution having semi-axes  $R, R, y$ , the corresponding strain energy should be written

$$\Delta g_s^\alpha = \{2\mu^\alpha(v^\beta - v^\alpha)^2/3v^\beta\} E(y/R), \tag{52.5}$$

and he obtained values for the function  $E(y/R)$  for some limiting cases. Since, in all the calculations,  $\Delta g_s^\alpha$  was shown to depend only on  $\mu^\alpha$  and not on the other elastic constant of the isotropic matrix, it seems reasonable to assume that this is true for all  $y/R$ .

When  $y/R = \infty$ , the ellipsoid becomes a cylinder (the mathematical model of a needle precipitate particle). This can be treated in the same way as the sphere; the displacements have the form  $\mathbf{w} = Ar$  inside the inclusion and  $\mathbf{w} = A(r_0^2/r^2)\mathbf{r}$  outside the inclusion, which is forced into a cylindrical hole of radius  $r_0$ . When the  $\beta$  inclusion is incompressible,  $A = 3\varepsilon/2$ , giving a fractional volume change of  $3\varepsilon$  and an increase of energy of

$$\Delta g_s^\alpha = \mu^\alpha(v^\beta - v^\alpha)^2/2v^\beta, \tag{52.6}$$

so that  $E(y/R) = \frac{3}{4}$ .

For a sphere,  $E(y/R) = 1$ . Nabarro showed that for a spheroid with  $y/R = 1 + \xi$ ,  $E(y/R)$  differs from unity only in terms of order  $\xi^2$  when  $\xi$  is small, so that  $E(y/R)$  is a slowly changing function near  $y/R = 1$ . Finally, when  $y/R \ll 1$ , the ellipsoid approximates to a thin plate or disc. Nabarro obtained an approximate solution for this case in which

$$E(y/R) \simeq 3\pi y/4R \quad (y/R \ll 1). \tag{52.7}$$

A smooth graph through these points leads to the variation of  $E(y/R)$  shown in Fig. 10.16. It must be emphasized that  $E(y/R)$  gives only the variation of the energy  $\Delta g_s^\alpha$ , and this is equivalent to  $\Delta g_s$  only when  $\Delta g_s^\beta \gg \Delta g_s^\alpha$ . Clearly this condition is satisfied for a plate nucleus, where the energy resides almost entirely in the matrix, and tends to zero as  $y/R$  decreases. Minimum strain energy is obtained by making the plate as thin as possible.

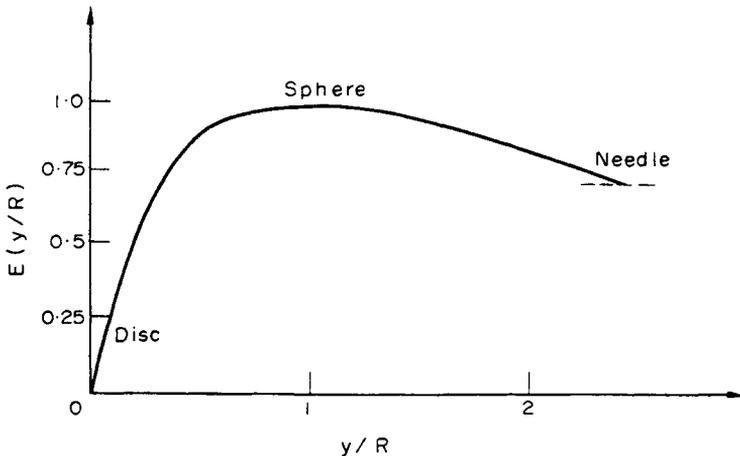


FIG. 10.16. To illustrate the variation of the strain energy of an incoherent nucleus with its shape (after Nabarro, 1940a).

As already stressed, the main limitation of the above results comes from the assumption of elastic isotropy. However, Kröner (1954) has made essentially equivalent calculations, making full allowance for crystal anisotropy. Although the details of the energy variation then depend on the individual elastic properties, the general shape of Fig. 10.16 is obtained for all curves of strain energy against  $y/R$ . Incoherent nuclei formed in a solid under constrained conditions will thus always have a shape approximating to a flat plate (oblate spheroid) if the volume change is appreciable. The smaller the  $y/R$  ratio, the smaller is the strain energy, and the larger is the surface energy for a given volume of nucleus. The most favourable nucleation path will utilize the shape (of finite  $y/R$ ) which minimizes the total free energy. Substituting expressions for volume and surface area into eqn. (52.1), we find

$$\Delta G = n \left[ g^\beta - g^\alpha + \frac{\pi}{4} \frac{y}{R} \mu^\alpha \frac{(v^\beta - v^\alpha)^2}{v^\beta} \right] + \pi^{1/3} (3v^\beta y / 4R)^{1/3} \times \left[ 2 + \frac{y^2}{R^2(1 - y^2/R^2)^{1/2}} \ln \left( \frac{1 + (1 - y^2/R^2)^{1/2}}{1 - (1 - y^2/R^2)^{1/2}} \right) \right] \sigma n^{2/3}, \quad (52.8)$$

where the axial ratio  $y/R$  specifies the shape of the ellipsoid. The saddle point in the energy field is determined by the simultaneous equations

$$(\partial \Delta G / \partial n)_{y/R} = 0, \quad [\partial \Delta G / \partial (y/R)]_n = 0, \quad (52.9)$$

which give the values  $n_c$ ,  $y_c$ ,  $R_c$  characteristic of the critical size nucleus of most favourable shape. The free energy of formation of this nucleus is obtained by substituting these values back into eqn. (52.8) and the homogeneous nucleation rate is then given by an equation of the form of (49.10).

We now consider the energy change when the  $\beta$  nucleus forms coherently within the  $\alpha$  phase. A general method of treating this problem in the approximation of isotropic elasticity has been given by Eshelby (1957). Suppose the free transformation  $\alpha \rightarrow \beta$  is represented by the homogeneous deformation  $\mathbf{y} = \mathbf{E}\mathbf{x}$  in an orthonormal coordinate system. The tensor  $\mathbf{E}$  specifies the deformation which transforms regions of  $\alpha$  into regions of  $\beta$ ; it gives the change of shape and volume undergone by these regions, but not necessarily the relations between the two lattices (see p. 322 and Part II, Chapters 21–22). Although the displacements  $\mathbf{w} = (\mathbf{E} - \mathbf{I})\mathbf{x}$  are not infinitesimal, we use them to define an elastic strain tensor

$$e_{ij}^T = \frac{1}{2} \left\{ \frac{\partial w_i}{\partial x_j} + \frac{\partial w_j}{\partial x_i} \right\} = \frac{1}{2} (E_{ij} + E_{ji}) - \delta_{ij}. \quad (52.10)$$

The suffix  $T$  is used to emphasize that the quantities  $e_{ij}^T$  are the strains in a free transformation which are to be distinguished from the strains  $e_{ij}^C$  which occur when the transformed region is constrained by the surrounding matrix. The main problem in considering the constrained transformation is the determination of  $e_{ij}^C$ .

Let  $X_{ij}^T$  be a stress field derived from  $e_{ij}^T$  by application of Hooke's law. Then from eqn. (11.26)

$$X_{ij}^T = \lambda \Delta^T \delta_{ij} + 2\mu e_{ij}^T. \quad (52.11)$$

The elastic constants used should be those appropriate to the  $\beta$  region, but for simplicity we shall assume that the  $\alpha$  and  $\beta$  phases have the same elastic properties for the remainder of this section. Now let the unit vector  $\mathbf{n}$  represent the outward normal at any point of the surface  $S$  separating the  $\beta$  particle and the  $\alpha$  matrix, and consider the sequence of operations detailed on p. 461. Since both the  $\beta$  and  $\alpha$  are unstressed at the end of stage 1, it follows that application of surface tractions  $-X_{ij}^T n_j$  over the surface of the  $\beta$  region will restore it elastically to its original size and shape, as required for stage 2. Thus after welding the surfaces together again, we are left with a layer of body force  $-X_{ij}^T n_j$  spread over the boundary surface, and this has to be annulled by a further distribution  $+X_{ij}^T n_j$ . We take the state of the assembly before adding this distribution (i.e. at the end of stage 3) as a state of zero displacement. The stress and strain in the matrix are then zero, and the  $\beta$  particle, although not stress-free, has the same external geometrical form as it had before transformation began. The displacements  $\mathbf{w}^C$  produced in matrix and  $\beta$  particle by the distribution  $X_{ij}^T n_j$  over  $S$  are thus the actual displacements involved in the whole process, and they define an elastic field given in all parts of the assembly by

$$e_{ij}^C = \frac{1}{2} \{ (\partial w_i^C / \partial x_j) + (\partial w_j^C / \partial x_i) \}.$$

The final stress in the matrix is written  $X_{ij}^C$ , and is derived from the elastic field  $e_{ij}^C$  by application of Hooke's law. However, the  $\beta$  nucleus had a stress field  $-X_{ij}^T$ , and a corresponding strain field  $-e_{ij}^T$  at the beginning of stage 4, so that its final stress field is

$$X_{ij}^I = X_{ij}^C - X_{ij}^T \tag{52.12}$$

and its final strain field is

$$e_{ij}^I = e_{ij}^C - e_{ij}^T.$$

If the  $\beta$  occupies a volume  $V$ , the elastic energy contained in it is

$$\frac{1}{2} \int_V (X_{ij}^I e_{ij}^I) dv = \frac{1}{2} \int_V X_{ij}^I (e_{ij}^C - e_{ij}^T) dv. \tag{52.13}$$

The elastic energy in the matrix is the integral  $\frac{1}{2} \int (X_{ij}^C e_{ij}^C) dv$  taken over the remaining volume of the assembly. However, this energy is also the work done on the surface  $S$  in setting up the elastic field, and may be written

$$-\frac{1}{2} \int_S (X_{ij}^C w_i^C n_j) dS = -\frac{1}{2} \int_S (X_{ij}^I w_i^C n_j) dS.$$

The negative sign corresponds to an outward direction of the positive normal to the element  $dS$ , and the two expressions are equal since the tractions and displacements are continuous across  $S$ . Use of Gauss's theorem, with the auxiliary conditions (eqns. (11.6))  $\partial X_{ij}^I / \partial x_j = 0$  and  $X_{ij}^I = X_{ji}^I$ , enables the second surface integral to be transformed into

$$-\frac{1}{2} \int_V (X_{ij}^I e_{ij}^C) dv. \tag{52.14}$$

Adding together (52.13) and (52.14), we find the total strain energy in the  $\beta$  nucleus and  $\alpha$  matrix is given by

$$-\frac{1}{2} \int_V X_{ij}^I e_{ij}^T dv. \tag{52.15}$$

$X_{ij}^T$  is defined by eqn. (52.11), and  $X_{ij}^C$  is given by a similar expression, so that  $X_{ij}^I$  may be written

$$X_{ij}^I = \lambda(\Delta^C - \Delta^T)\delta_{ij} + 2\mu(e_{ij}^C - e_{ij}^T). \quad (52.16)$$

The quantities  $e_{ij}^T$  being specified for a given transformation, the strain energy can be calculated from eqns. (52.15) and (52.16) if  $w_i^C$  and hence  $e_{ij}^C$  can be found. Eshelby showed that a general expression for  $w_i^C$  may be written in terms of the derivatives of two functions which are respectively the Newtonian potential and the biharmonic potential of attracting matter filling the volume  $V$  bounded by  $S$ . We confine ourselves to an inclusion of ellipsoidal shape, where explicit solutions were possible. The strains  $e_{ij}^C$  are then uniform, and depend only on the shape (this result remains true in an anisotropic medium), so that the strain energy per atom may be obtained from (52.15) as

$$\Delta g_s = -\frac{1}{2} X_{ij}^I e_{ij}^T v^{\beta}. \quad (52.17)$$

It is convenient to use the axes of the ellipsoid as coordinate axes, and Eshelby proved that the relation between the constrained and stress-free strains can be written

$$e_{il}^C = s'_{ilmn} e_{mn}^T, \quad (52.18)$$

where the quantities  $s'_{ilmn}$  are determined by the values of certain elliptic integrals. Coefficients linking shears to extensions ( $s'_{1123}$ ,  $s'_{2221}$ , etc.) are zero, as also are those connecting two shears ( $s'_{1223}$ , etc.), but  $s'_{1122} \neq s'_{2211}$ , etc. We consider only the simple model in which the ellipsoid is a flat oblate spheroid having axes  $R$ ,  $R$ ,  $y$  with  $R \gg y$ . This corresponds to the model of the flat plate precipitate used by Nabarro for incoherent precipitation. The elliptic integrals then reduce to simple forms, and the following results for  $s'_{ilmn}$  may be derived (the  $x_3$  axis is directed along  $y$ , and  $\nu$  is Poisson's ratio).

$$\left. \begin{aligned} s'_{1111} = s'_{2222} &= \frac{(13-8\nu)}{32(1-\nu)} \pi \frac{y}{R}, \\ s'_{1122} = s'_{2211} &= \frac{(8\nu-1)}{32(1-\nu)} \pi \frac{y}{R}, \\ s'_{1133} = s'_{2233} &= -\frac{(1-2\nu)}{8(1-\nu)} \pi \frac{y}{R}, \\ s'_{3333} &= 1 - \frac{(1-2\nu)}{4(1-\nu)} \pi \frac{y}{R}, \\ s'_{3311} = s'_{3322} &= \frac{\nu}{(1-\nu)} - \frac{(1+4\nu)}{8(1-\nu)} \pi \frac{y}{R}, \\ s'_{1313} = s'_{2332} &= \frac{1}{2} - \frac{(2-\nu)}{8(1-\nu)} \pi \frac{y}{R}, \\ s'_{1212} &= \frac{(7-8\nu)}{32(1-\nu)} \pi \frac{y}{R}. \end{aligned} \right\} \quad (52.19)$$

We are now able to calculate  $\Delta g_s$  for various possibilities of physical interest. For example, if the transformation is a pure dilatation, we have from (52.18)

$$\Delta C = (2\Delta^T/3)(s'_{111} + s'_{122} + s'_{133} + s'_{331} + \frac{1}{2}s'_{333}) = \frac{(1+\nu)}{3(1-\nu)}\Delta^T.$$

On substituting into (52.16) and eliminating  $\lambda$ , we find

$$\Delta g_s = \frac{2}{9} \frac{(1+\nu)}{(1-\nu)} \mu (\Delta^T)^2 \nu^\beta. \tag{52.20}$$

This equation does not contain  $y/R$ , and is in fact identical with the energy of the misfitting sphere model (25.16) or (52.2) if the  $\beta$  and  $\alpha$  regions have the same elastic constants. It may be proved quite generally that (52.20) always gives the strain energy for a uniform dilatation, whatever the shape of the particle, a result due originally to Crum and quoted by Nabarro (1940a). Since the strain energy in a rigid matrix would be  $\Delta g_s^\beta = \frac{1}{2}K(\Delta^T)^2 \nu^\beta$ , the accommodation factor is

$$1 - C_6 = \frac{4}{9} \frac{(1+\nu)}{(1-\nu)} \frac{\mu}{K} = \frac{2}{3} \frac{(1-2\nu)}{(1-\nu)} \approx \frac{1}{3}. \tag{52.21}$$

This accommodation factor is identical with that for the incoherent spherical precipitate except that we are now ignoring the differences in elastic properties. In contrast to the behaviour of the incoherent nucleus, however, the accommodation factor for a coherent nucleus stays constant for all shapes if the transformation is a uniform dilatation.

If we now consider other types of transformation, it is physically obvious that good accommodation (low energy) is possible for displacements perpendicular to the face of a flat plate, but not for displacements parallel to the plate. In the most general case, the transformation has principal axes parallel to those of the ellipsoid, but the changes in length are unequal. The transformation is specified by

$$e_{11}^T = \varepsilon_1, \quad e_{22}^T = \varepsilon_2, \quad e_{33}^T = \varepsilon_3,$$

and some rather lengthy algebra leads to the strain energy

$$\begin{aligned} \Delta g_s = & \frac{\mu\nu^\beta}{(1-\nu)} \{ \varepsilon_1^2 + \varepsilon_2^2 + 2\nu\varepsilon_1\varepsilon_2 \} - \frac{\mu\nu^\beta}{32(1-\nu)} \pi \frac{y}{R} \{ 13(\varepsilon_1^2 + \varepsilon_2^2) + 2(16\nu - 1)\varepsilon_1\varepsilon_2 \\ & - 8(1+2\nu)(\varepsilon_1 + \varepsilon_2)\varepsilon_3 - 8\varepsilon_3^2 \}. \end{aligned} \tag{52.22}$$

The limiting value of  $\Delta g_s$  as  $c/a \rightarrow 0$  is thus finite, except when  $\varepsilon_1 = \varepsilon_2 = 0$ , and only in this case is complete accommodation possible. On the other hand, when  $\varepsilon_3 = 0$ , the energy is not greatly reduced below the value  $\Delta g_s^\beta = [\mu/(1-2\nu)] \{ (1-\nu)(\varepsilon_1^2 + \varepsilon_2^2) + 2\nu\varepsilon_1\varepsilon_2 \}$  which would result from transformation in a rigid matrix. Two special cases are  $\varepsilon_2 = 0$ , giving a transformation in which the two planes are constrained into coherency by stretching along one atomic direction, and  $\varepsilon_1 = \varepsilon_2$ , giving a transformation in which the two phases have

the same arrangement of atoms on some plane, but slightly different interatomic distances. In the first case, the accommodation factor is

$$1 - C_6 = \frac{(1 - 2\nu)}{(1 - \nu)^2} \approx \frac{3}{4} \quad (52.23)$$

and in the second case it is

$$1 - C_6 = \frac{(1 + \nu)(1 - 2\nu)}{(1 - \nu)} \approx \frac{2}{3}. \quad (52.24)$$

Thus for these two types of coherency, the strain energy resides mainly in the nucleus, and change of shape is rather ineffective in reducing the energy. Nabarro (1940a) gave a rough calculation for an anisotropic (cubic) crystal to show that the strain energy of a coherent flat precipitate with  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3$  is one-sixth that of a spherical precipitate in which  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon$ . Equation (52.22) shows that in the limit of very small  $y/R$  these two energies are identical in the isotropic approximation, as would also be expected from the result that the strain energy of a uniform dilatation is independent of shape.

It is important to note that the strain energy of a coherent nucleus can be very large in comparison with the energies which can usually be stored in a metal by deformation. If the transformation involves a 10% volume change, for example, the strain energy of a uniform dilatation with  $\mu = 8 \cdot 10^{10} \text{ N m}^{-2}$  and  $\nu = \frac{1}{4}$  is given by (52.20) as  $3 \cdot 10^8 \text{ J m}^{-3}$ , or (say) 400–500 cal mole<sup>-1</sup>. Strain energies of the same order are obtained for any transformation in which there is a misfit of a few per cent in any direction in the plane of the flat ellipsoid representing the nucleus. These energies are comparable with the chemical free energies which provide the driving force for transformation, and it follows that in all such cases the nucleus must break away to relieve its strain energy at an early stage in transformation. As already emphasized, the existence of such high strain energies and corresponding high internal stresses is only possible because of the small volumes of the particles concerned. Very small coherent particles are important in precipitation reactions in the solid state (Part II, Chapter 16), and they can harden the material very considerably. Because of the magnitude of the strain energy, however, reasonably large precipitates can remain coherent only if macroscopic length changes in the interface are less than about 1%.

If in (52.22) we now put  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \Delta^T/3$ , we recover eqn. (52.20). A slightly different possibility is the combination of a uniform dilation  $\Delta_1$  with a uniaxial strain normal to the plate. Substituting  $\varepsilon_1 = \varepsilon_2 = \Delta_1/3$ ,  $\varepsilon_3 = \xi + \Delta_1/3$ , we find the strain energy

$$\Delta g_s = \frac{\mu v^\beta}{(1 - \nu)} \left\{ \frac{2}{9} (1 + \nu) \Delta_1^2 + \frac{\pi}{4} \frac{y}{R} \xi^2 + \frac{(1 + \nu)}{3} \pi \frac{y}{R} \Delta_1 \xi \right\}. \quad (52.25)$$

This expression is important in the theory of martensite nucleation (Christian, 1958).

In the above examples, the principal directions of  $e_{ij}^T$  coincided with the axes of the ellipsoid, and the two phases were constrained to coherency by expansions or contractions parallel to these directions. Now consider the strain energy when the  $\beta$  phase is formed from the  $\alpha$  phase by a shear transformation in which  $e_{13}^T = e_{31}^T = s/2$  are the only non-zero components of  $e_{ij}^T$ . This gives

$$\Delta g_s = \frac{1}{2} \mu v^\beta (1 - 2s'_{1313}) s^2 \quad (52.26)$$

and the accommodation factor is

$$1 - C_6 = 1 - 2s'_{1313} = \frac{(2 - \nu)}{4(1 - \nu)} \pi \frac{y}{R} \approx 2 \frac{y}{R}, \quad (52.27)$$

so that the strain energy is small when  $y/R$  is very small. For reasonably small  $y/R$ , there is good accommodation, and most of the strain energy is contained in the surrounding matrix. Since the coefficients  $s'_{ilmn}$  linking shear and non-shear components of  $e_{ij}^C$  and  $e_{mn}^T$  are zero, the strain energy in a transformation in which the change is an invariant plane strain is simply the sum of (52.26) and (52.25) with  $\Delta = 0$ . This energy tends to zero as  $y/R$  decreases, so that the shape change of an invariant plane strain can be accommodated in the matrix with very small energy if  $y/R$  is small enough, a result which is, perhaps, physically obvious. Note that if the operative shear components are  $e_{12}^T = e_{21}^T$ , the corresponding accommodation factor  $1 - 2s'_{1212}$  approaches unity as  $y/R$  decreases. There is thus no accommodation when the shear tends to deform the plate in its own plane.

The application of these results to the theory of nucleation may now be made in the same way as before. Suppose, for example, that a coherent nucleus forms in such a way that two crystallographic planes from the two phases are parallel and have similar atomic arrangements, but different atomic spacing. The simplest case is when the atoms are in disregistry along one atomic direction only by an amount  $\varepsilon_1$ , so that the strain energy when the phases remain completely coherent is

$$\Delta g_s = \mu v^\beta \varepsilon_1^2 / (1 - \nu). \quad (52.28)$$

The rate of nucleation is then given by an equation similar to (52.8), with the above expression for  $\Delta g_s$ , replacing the strain energy of the incoherent particle. Obviously,  $\Delta g_s$  is much larger for coherent precipitation, but the surface energy  $\sigma$  is correspondingly smaller. When the nucleus is small, it is probable that the surface term is predominant, and coherent nucleation is favoured. Nevertheless, the condition that the most favourable nucleation path be chosen also requires consideration of the intermediate case of semi-coherent nucleation. As a model for this, we may assume that the strain in the  $\beta$  lattice in the  $x_1$  direction is  $\varepsilon'_1$  instead of  $\varepsilon_1$ , the remaining difference in length of  $\varepsilon_1 - \varepsilon'_1$  being obtained by an array of dislocations of density proportional to  $\varepsilon_1 - \varepsilon'_1$ . This introduces an extra surface energy term, which may be taken to be proportional to the dislocation density in a first approximation, so that for semi-coherent nucleation, the surface free energy per unit area becomes

$$\sigma = \sigma' + C_{11}(\varepsilon_1 - \varepsilon'_1). \quad (52.29)$$

At the same time, the accommodation strain energy is reduced to

$$\Delta g_s = \mu v^\beta (\varepsilon'_1)^2 / (1 - \nu). \quad (52.30)$$

It is clear that a semi-coherent nucleus will form if it is cheaper in energy to accommodate some of the elastic misfit by the introduction of dislocations into the surface region. Equation (52.8) gives the energy of this semi-coherent nucleus if (52.30) is used in place of (52.5) for the strain energy and (52.29) is substituted for  $\sigma$ . The saddle point free energy change

is now defined so that  $\Delta G$  should be a maximum with respect to changes in size and a minimum with respect to changes in shape and degree of coherency. This gives the auxiliary conditions

$$\left. \begin{aligned} [\partial\Delta G/\partial n]_{y/R, \varepsilon_i} &= 0, \\ [\partial\Delta G/\partial(y/R)]_{n, \varepsilon_i} &= 0, \\ [\partial\Delta G/\partial\varepsilon'_i]_{n, y/R} &= 0, \end{aligned} \right\} \quad (52.31)$$

which together define the combination of size, shape, and degree of coherency for the critical nucleus.

Essentially equivalent considerations are involved when the misfit is not confined to a single atomic direction. In principle, the free energy for the most general case of a semi-coherent precipitate will be a function of  $n$ ,  $y/R$ , and three parameters  $\varepsilon'_1$ ,  $\varepsilon'_2$ ,  $\varepsilon'_3$  representing the degree of coherency. Very detailed data would be required to establish the favoured mode of nucleation in any particular case, since a knowledge of the anisotropy of both elastic properties and surface energy would be needed.

We have already emphasized that real nucleation in the solid state, as in other assemblies of atoms, is nearly always heterogeneous. It is, indeed, rather more difficult to obtain homogeneous solid nucleation, since the production of crystals nearly free from defects is harder than the production of assemblies nearly free from impurities; some results have, however, been achieved by the small particle method. The effectiveness of nucleation catalysts in reducing the surface energy term has already been fully discussed, but there are extra considerations in the solid state. As mentioned on p. 463, certain preferred sites, in particular grain boundaries, may be able to accommodate the shape and volume changes by flow at much lower temperatures than is possible in a region of nearly perfect lattice. The reduction of the  $\Delta g_s$  term to nearly zero then gives a lower critical free energy of nucleation, and a correspondingly more rapid nucleation rate.

On p. 463, we gave reasons for believing that single dislocations or Frank-Read sources would be ineffective in causing flow of this kind. Nevertheless, a large number of experimental observations show conclusively that dislocations do act as preferred sites for the formation of nuclei. Most results are for precipitation from solid solutions, where a composition change has to be produced, and dislocation sites are presumably favoured because the Cottrell atmosphere effect makes them suitable centres for segregation of solute atoms. However, there should be an analogous effect even in a pure component; solute atoms are attracted to a dislocation because they can lower its strain energy, and the same result can be obtained if the atoms around the dislocation line are rearranged into some new stable pattern.

A theory of nucleation on dislocations has been given by Cahn (1957). He assumes that the nucleus lies along the dislocation and has a circular section perpendicular to the dislocation line. The radius of the section is not constant, but varies with distance along the line, so that the longitudinal section if the nucleus is approximately as shown in Fig. 10.17. In addition to the usual volume and surface energy terms in the expression for the energy of formation of a nucleus of given size, there is a term representing the strain energy of the dislocation in the region now occupied by the new phase. In effect, the atomic rearrangement

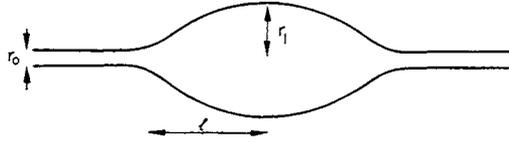


FIG. 10.17. The formation of a nucleus on a dislocation line (after Cahn, 1957).

within the nucleus is assumed to destroy that part of the elastic energy of the dislocation located in the volume of the nucleus, and the energy thus gained is available to help the nucleation process.

Suppose a small length of the nucleus is effectively a cylinder of radius  $r$  centred on the dislocation. The change in free energy per unit length when the nucleus is formed is given by

$$\Delta G = \pi r^2(g^\beta - g^\alpha)/v^\beta + C_{12} - \frac{1}{2}Bb \ln r + 2\pi\sigma r, \tag{52.32}$$

where the first term is the (negative) volume free energy change, the term  $C_{12} - \frac{1}{2}Bb \ln r$  (see eqns. (30.16) and (30.18)) represents the dislocation energy within the radius  $r$ , and the last term is the positive surface energy. The constant  $B$  varies between values  $B_e$  for an edge dislocation, and  $B_s$  for a screw dislocation. If the transformation in a region of good  $\alpha$  crystal involves an appreciable strain energy  $\Delta g_s$ , this should be added to the first term, as before.

The important difference between eqn. (52.32) and previous expressions for the change in energy on forming a nucleus is that there are two negative terms, the chemical free energy change, which becomes predominant at large values of  $r$ , and the dislocation strain energy released, which is most important at small values of  $r$ . At intermediate radii, the positive surface energy may force an increase in  $\Delta G$ , but this does not necessarily happen. Differentiating the equation, and equating  $\partial\Delta G/\partial r = 0$  gives

$$r = \frac{\sigma v^\beta}{2(g^\alpha - g^\beta)} \left[ 1 \pm \left\{ 1 - \frac{(g^\alpha - g^\beta)Bb}{\pi v^\beta \sigma^2} \right\}^{1/2} \right]. \tag{52.33}$$

The behaviour thus depends on whether the quantity  $\alpha^D = (g^\alpha - g^\beta)Bb/\pi v^\beta \sigma^2$  is greater or less than one. If  $\alpha^D > 1$ , there are no turning points in the  $\Delta G - r$  relation, and the energy of the whole nucleus (whatever its exact shape) decreases continually as  $r$  increases. If  $\alpha^D < 1$ , there is a minimum free energy at a value  $r = r_0$  corresponding to the negative sign in eqn. (52.33), followed by a maximum at a value of  $r$  corresponding to the positive sign. These two possibilities are shown in Fig. 10.18. As  $\alpha^D$  increases from 0 to 1, the value of  $r_0$  changes only by a factor of two, from  $Bb/4\pi\sigma$  to  $Bb/2\pi\sigma$ .

When the chemical free energy and dislocation energy factors are sufficiently large in comparison with the surface energy, there is thus no energy barrier to nucleation on dislocations, and transformation rates will be governed only by growth conditions. When the free energy curve has the alternative form  $A$  (Fig. 10.18), there will be a sub-critical metastable cylinder of the  $\beta$  phase surrounding the dislocation line to the radius of the minimum in the curve, and this is roughly analogous to the Cottrell atmosphere of solute atoms in a segregation problem. When the two phases are in equilibrium,  $\alpha^D = 0$ , and the maximum

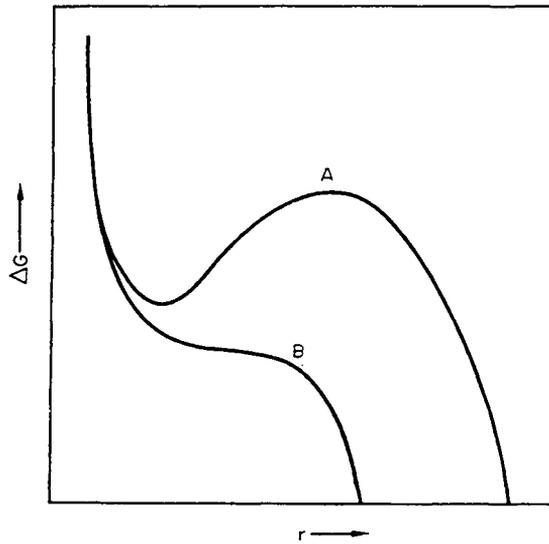


FIG. 10.18. Schematic forms for the variation of the free energy of formation of a dislocation nucleus with its radius (after Cahn, 1957). Curve *A*,  $\alpha^D < 1$ . Curve *B*,  $\alpha^D > 1$ .

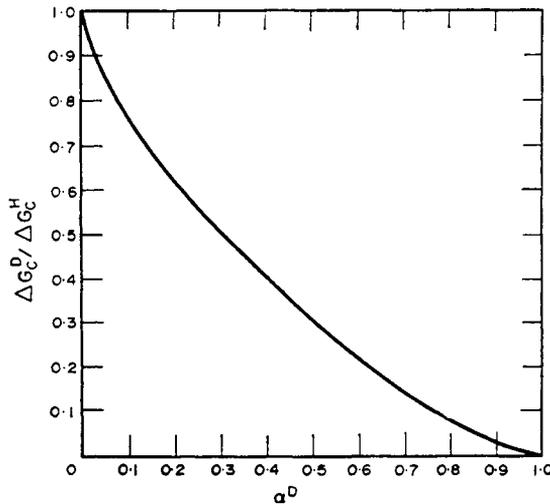


FIG. 10.19. The ratio of the free energy required to form a nucleus on a dislocation to that required to form a homogeneous spherical nucleus is plotted as a function of the parameter  $\alpha^D$  (after Cahn, 1957).

in the free energy change of eqn. (52.32) is infinite, as for homogeneous nucleation. As the  $\beta$  phase becomes relatively more stable,  $\alpha^D$  increases until the nucleation rate becomes appreciable; it is shown below that this usually requires  $\alpha^D$  to be in the range 0.4–0.7. The situation in curve *B* (Fig. 10.18) will only be obtained if the conditions are altered so rapidly that  $\alpha^D$  exceeds unity before appreciable transformation has occurred.

Equation (52.32) and the curves of Fig. 10.18 refer only to the energy per unit length of a small element of the nucleus. The energy of the whole nucleus will be obtained by integrating along the effective length of the nucleus, shown in Fig. 10.17. To find the nucleation rate, the combination of size and shape which give the saddle point energy change for the whole nucleus is required. Cahn solved this problem by applying the appropriate boundary conditions; the critical nucleus (Fig. 10.17) is defined in terms of a maximum radius  $r_1$  and a measure of its effective length  $l$ . His results are shown in Fig. 10.19, in which the ratio of the critical free energy of nucleation to the corresponding free energy of a homogeneous spherical nucleus of critical size is plotted against the function  $\alpha^D$ .

Let  $\Delta G_c^D$  be the critical free energy of nucleus formation on a dislocation. Then the total nucleation rate will be given by eqn. (49.10), replacing  $\Delta G_c$  by  $\Delta G_c^D$  and  $N$  by  $N^D$ , the total number of atoms on dislocation lines. More usefully, the nucleation rate per unit volume,  $vI^D$ , will be given by

$$\begin{aligned} vI^D &= vN^D (kT/h) \exp\{-(\Delta G_c^D + \Delta_a g^*)/kT\} \\ &\approx (vN)^{1/3} vL^D (kT/h) \exp\{-(\Delta G_c^D + \Delta_a g^*)/kT\}, \end{aligned} \tag{52.34}$$

where  $vN^D$  is the number of atoms per unit volume which are on dislocation lines, and  $vL^D$  is the dislocation density, that is the length of dislocation line per unit volume. The second expression for  $vI^D$  is correct only if dislocations are regarded essentially as lines one atom diameter in cross-section; more realistically, it should be multiplied by a small numerical factor, say about five. As always, the pre-exponential term in the nucleation rate is ill-defined, but relatively unimportant to one or two orders of magnitude. We see that the volume nucleation rate is linearly proportional to the dislocation density, and the condition for the nucleation rate to become appreciable ( $vI^D \sim 10^6 \text{ m}^{-3} \text{ s}^{-1}$ ) is, for  $vL^D$  in  $\text{m}^{-2}$ ,

$$\Delta G_c^D + \Delta_a g^* = kT(39 + \ln vL^D). \tag{52.35}$$

The function  $f(\alpha^D)$  shown in Fig. 10.19 gives the ratio of  $\Delta G_c^D$  to  $\Delta G_c$  of eqn. (46.19), so that in terms of this function, condition (52.35) becomes

$$\frac{1}{(\alpha^D)^2} f(\alpha^D) = \frac{3\pi\sigma kT}{16(Bb)^2} \left[ 39 + \ln vL^D - \frac{\Delta_a g^*}{kT} \right]. \tag{52.36}$$

With typical values of  $vL^D = 10^{12} \text{ m}^{-2}$ ,  $\Delta_a g^* \approx 10kT$ , this gives  $(1/\alpha^D)^2 f(\alpha^D) \approx 34\sigma kT/(Bb)^2$ , and the numerical factor is rather insensitive to the actual value of  $vL^D$  and  $\Delta_a g^*/kT$ . A graph of  $(1/\alpha^D)^2 f(\alpha^D)$  against  $\alpha^D$  thus gives the relation between  $34\sigma kT/(Bb)^2$  and  $\alpha^D$  needed to ensure rapid nucleation. Reasonable values of  $B$  and  $\sigma$  give  $34\sigma kT/(Bb)^2 \approx 1$  at 1000 K, so that for most transformations of practical interest  $(1/\alpha^D)^2 f(\alpha^D)$  probably falls in the range 0.5–5, corresponding to  $\alpha^D$  values between 0.4 and 0.7. From the value of  $\alpha^D$ , the required driving force  $g^\alpha - g^\beta$  may be estimated.

It is interesting to compare the nucleation rate on dislocations, given by (52.34), with the corresponding homogeneous nucleation rate. Using  $\sigma = 0.2 \text{ J m}^{-2}$  (200 ergs  $\text{cm}^{-2}$ ), and a value of  $g^\alpha - g^\beta$  corresponding to  $\alpha \approx 0.6$ , the homogeneous nucleation rate given by eqn. (49.10) may be estimated at  $10^{-64} \text{ m}^{-3} \text{ s}^{-1}$ , whereas the nucleation rate on dislocations

under these conditions is  $10^{+14} \text{ m}^{-3} \text{ s}^{-1}$ . Nucleation would thus be entirely confined to dislocations if the assumptions of the model are correct. The assumptions are physically self-consistent, since  $r_0$ ,  $r_1$  have the very reasonable values 0.2 nm, 1 nm (2 Å, 10 Å) when  $\alpha^D \simeq 0.6$ . Since the ratio  $\Delta G_c^D/\Delta G_c^H$  decreases as  $\alpha^D$  increases, nucleation is easier on dislocations with large Burgers vector. Also, since  $\alpha^D$  is proportional to  $g^\alpha - g^\beta$ , the rate of change of dislocation nucleation rate with increasing  $g^\alpha - g^\beta$  (i.e. with temperature or supersaturation) must be even more rapid than that of the homogeneous nucleation rate.

In Cahn's calculation, the assumption that the whole of the strain energy of the dislocation within the volume occupied by the nucleus can be relaxed to zero requires that the nucleus be incoherent. For a coherent nucleus forming on or near dislocations, this assumption cannot be made; instead it is necessary to calculate the elastic interaction energy between the nucleus and the matrix. A calculation of this type for nuclei in which  $e_{ij}^T$  corresponds to a pure dilatation and the dislocation is edge in character has been made by Dollins (1970) and improved by Barnett (1971). The problem for a spherical nucleus with its centre at  $r$ ,  $\theta$ ,  $z$  with respect to the dislocation line is clearly identical with the treatment given in Section 30; Barnett used the correct (Eshelby) formulation and obtained for the interaction energy in the case when nucleus and matrix have the same elastic properties

$$W_i = \frac{2}{3} B_e (1 + \nu) \Delta^T v \sin \theta / r, \quad (52.37)$$

which is equivalent to (30.41) with the misfit volume replaced by  $\Delta^T v$ , where  $\Delta^T$  is the cubical dilatation and  $v$  is the unconstrained nucleus volume. This same equation also represents the interaction energy for a coherent nucleus in the form of an ellipsoid of revolution if the stress-free strain is a pure dilatation, and it may therefore also be applied to a disc-like precipitate.

When the nucleus has different elastic properties from the matrix, the interaction energy is more difficult to estimate. The energy may be divided into two parts

$$W_i = W_i^{(1)} + W_i^{(2)}, \quad (52.38)$$

where  $W_i^{(1)}$  results from the interaction of the stress field of the dislocation with the strain field of the nucleus, as above, whereas  $W_i^{(2)}$  is the so-called modulus effect in which, for example, a region of higher elastic modulus increases the self-energy of the dislocation.

In the case of a sphere, an exact calculation is possible and yields

$$W_i^{(1)} = (\mu^\alpha b / \pi) \{3K^\beta / (3K^\beta + 4\mu^\alpha)\} \Delta^T v \sin \theta / r \quad (52.39)$$

where  $\mu^\alpha$ ,  $K^\beta$  are elastic stiffness of the matrix and the nucleus respectively. This equation is equivalent to eqns. (30.42) and (25.8) with  $\Delta v_{AB}$  in the latter replaced by  $\Delta^T v$ . It differs from (52.37) only by the factor

$$\{\kappa^\alpha + (K^\alpha / K^\beta) (1 - \kappa^\alpha)\}^{-1} \quad (52.40)$$

where

$$\kappa^\alpha = (1 + \nu^\alpha) / 3(1 - \nu^\alpha)$$

When  $K^\alpha = K^\beta$ , the second bracketed term in (52.39) reduces to  $\kappa^\alpha$ , and (52.39) becomes equivalent to (52.37).

The modulus effect was estimated by Dollins and Barnett by making the assumption that the strain field of the dislocation is not changed by the nucleus. Barnett gives for the ellipsoid of revolution of semi-axes  $R, R, y$ .

$$W_i^{(2)} = \frac{(\mu^\beta - \mu^\alpha)b^2y}{2\pi(1-\nu^\alpha)} \{1 - (\beta^2 - 1)^{1/2} \tan^{-1}(\beta^2 - 1)^{-1/2}\} \quad (52.41)$$

where  $\beta = r/R > 1$ . However, the approximation used to calculate  $W_i^{(2)}$  is generally not a good one, since the strain field of the dislocation inside the nucleus is changed appreciably from its matrix value when the nucleus has different elastic properties.

In general  $W_i^{(1)}$  can always be made negative by suitable choice of  $\sin \theta$ ; i.e. the nucleus can form on the compression or tension side of the edge dislocation, depending on the sign of the volume change. The sign of  $W_i^{(2)}$  depends on whether  $\mu^\beta$  is greater or smaller than  $\mu^\alpha$ . Nucleation will occur preferentially in the vicinity of the dislocation provided  $W_i$  is negative, and the net driving force is then increased. In contrast to Cahn's theory, however, the nucleation barrier  $\Delta G_c$  may be reduced but not eliminated by this type of dislocation-catalysed nucleation.

Since a screw dislocation has no hydrostatic stress field in first order approximation, nucleation of coherent precipitates near screws will not be expected when, as assumed above, the shape change is a pure dilatation. In the case of martensitic transformations, however, the stress free strains correspond mainly to shears, and preferential regions for nucleation may thus be found near to screw dislocations. A treatment of this problem would be similar to that given for the interaction of an interstitial defect with a screw dislocation; see eqn. (30.43).

Gomez-Ramirez and Pound (1973) have developed a model for nucleation on dislocations which is in some respects intermediate between Cahn's model and the model just described. It is assumed that the nucleus will form along the dislocation because of the importance of the core energy, but an Eshelby-type calculation is used to estimate the reduction in strain energy. Instead of Cahn's assumption that the elastic energy within the nucleus is released, it is supposed that for an incoherent nucleus the dislocation along the centre of the nucleus is replaced by a distribution of infinitesimal dislocations over its surface. The energy change resulting from formation of the nucleus along the dislocation rather than in the defect-free lattice then contains three terms, namely:

- (1) the core energy, assumed to be uniformly distributed over the volume of the core;
- (2) the elastic interaction energy, calculated as above, except that since  $\sin \theta/r$  varies rapidly it is necessary to take the integral of  $X_{ij}^d e_{ij}^T$  over the volume of the embryo;
- (3) the change in elastic energy caused by "spreading" the Burgers vector into the interface.

For the particular case of a screw dislocation and a cylindrical embryo, the elastic energy of the smeared out dislocations is calculated to be fortuitously equal to the elastic energy of the original dislocation outside the embryo, and this is assumed to be generally a good estimate. Thus the contribution (3) is taken to be the negative of the elastic energy of the

dislocation, integrated over the volume of the embryo, but excluding the core region. Note that this corresponds fairly closely to Cahn's assumption.

With this model it was found impossible to minimize  $\Delta G$  analytically, but numerical calculation showed the energy barrier to be only slightly shape dependent. Numerical calculations were therefore carried out for assumed smooth shapes and isotropic surface free energies. Embryo shapes along screw dislocations are predicted to be similar to prolate spheroids with eccentricity  $\leq 0.83$ , whilst similar closed shapes but with heart-shaped cross-sections to allow for the asymmetry of the dislocation field are proposed for edge dislocations. Metastable embryos form along dislocations in both stable and metastable  $\alpha$  regions if the volume misfit is small and the shear modulus is sufficiently high; but the automatic formation of metastable tubes of  $\beta$  phase, as in Cahn's theory, is not predicted. (This difference presumably arises from different estimates of the core energy density, or of the effective cut off radius in the elastic calculation.)

As would be expected for a model in which the strain field of the embryo is a pure dilatation, the critical free energy of formation on an edge dislocation  $\Delta G_c^{De}$ , is found to be smaller than the corresponding energy on a screw dislocation  $\Delta G_c^{Ds}$ . However, in all cases

$$\Delta G_c^H - \Delta G_c^{Ds} > \Delta G_c^{Ds} - \Delta G_c^{De},$$

and with reasonable values of dislocation density the nucleation rates per unit volume were greatest for edge and smallest for homogeneous nucleation. Thus for the transformation in pure iron at a driving force of  $\sim 4.8 \times 10^8 \text{ J m}^{-3}$  ( $115 \text{ cal cm}^{-3}$ ) with an assumed interfacial free energy of  $0.24 \text{ J m}^{-2}$ , the authors calculate  $\Delta G_c^H$ ,  $\Delta G_c^{Ds}$ , and  $\Delta G_c^{De}$  as 7.1, 1.6, and 0.9 eV respectively. If the dislocation density is  $\sim 10^{10} \text{ m}^{-2}$ , the corresponding nucleation rates are  $10^{5.5}$ ,  $10^{24.5}$ , and  $10^{26.8} \text{ m}^{-3} \text{ s}^{-1}$  respectively. It follows that nuclei will form first on dislocations of predominantly edge character, and that these may effect the whole transformation at low driving forces; at high driving forces, however, nucleation should take place on all types of dislocations and possibly also homogeneously in the matrix.

#### REFERENCES

- ABRAHAM, F. F. (1969) *J. Chem. Phys.* **51**, 1632.  
 BARNETT, D. M. (1971) *Scripta metall.* **5**, 261.  
 BECKER, R. (1940) *Proc. Phys. Soc.* **52**, 71.  
 BECKER, R. and DÖRING, W. (1935) *Ann. Phys.* **24**, 719.  
 BENSON, G. C. and SHUTTLEWORTH, R. (1951) *J. Chem. Phys.* **19**, 130.  
 BUFF, F. P. (1951) *J. Chem. Phys.* **19**, 1591.  
 BURTON, J. J. (1973) *Acta metall.* **21**, 1225.  
 CAHN, J. W. (1956) *Acta metall.* **4**, 449; (1957) *Ibid.* **5**, 168.  
 CARSLAW, H. S. and JAEGER, J. C. (1947) *Conduction of Heat in Solids*, p. 82, Clarendon Press, Oxford.  
 CHRISTIAN, J. W. (1958) *Acta metall.* **6**, 377.  
 CLEMM, P. J. and FISHER, J. C. (1955) *Acta metall.* **3**, 70.  
 COURANT, R. and HILBERT, D. (1953) *Methods of Mathematical Physics*, p. 76, Interscience, New York.  
 DAWSON, D. B., WILSON, E. J., HILL, P. G., and RUSSELL, K. C. (1969) *J. Chem. Phys.* **51**, 5389.  
 DOLLINS, C. C. (1970) *Acta metall.* **18**, 1209.  
 DUNNING, W. J. (1965) *Symposium on Nucleation*, p. 1, Case Institute of Technology, Cleveland, Ohio; (1969) *Nucleation*, p. 1, Dekker, New York.  
 ESHELBY, J. D. (1957) *Proc. R. Soc. A*, **241**, 376; (1961) *Prog. Solid Mech.* **2**, 89.

- FARKAS, L. (1927) *Z. phys. Chem.* **125**, 239.
- FEDER, J., RUSSELL, K. C., LOTHE, J., and POUND, G. M. (1966) *Adv. Phys.* **15**, 111.
- FISHER, J. C., HOLLOMON, J. H., and TURNBULL, D. (1948) *J. Appl. Phys.* **19**, 775.
- FRENKEL, J. (1939) *J. Chem. Phys.* **1**, 200, 538; (1946) *Kinetic Theory of Liquids*, Oxford University Press.
- GOMEZ-RAMIREZ, R. and POUND, G. M. (1973) *Metall. Trans.* **4**, 1563.
- HILLIG, W. B. (1962) *Symposium on Nucleation and Crystallisation in Glasses and Melts*, p. 77, American Ceramic Society.
- HIRTH, J. P. and POUND, G. M. (1963) *Prog. Mater. Sci.* **11**, 1.
- HOLLOMON, J. H. and TURNBULL, D. (1953) *Prog. Metal Phys.* **4**, 333.
- JAEGER, H. L., WILSON, E. J., HILL, P. G., and RUSSELL, K. C. (1969) *J. Chem. Phys.* **51**, 5380.
- JOHNSON, W. C., WHITE, C. L., MARTH, P. E., RUF, P. K., TUOMINEN, S. M., WADE, K. D., RUSSELL, K. C. and AARONSON, H. I. (1975) *Metall. Trans.* **6A**, 911.
- KANTROWITZ, A. (1951) *J. Chem. Phys.* **19**, 1097.
- KIRKWOOD, J. G. and BUFF, F. P. (1949) *J. Chem. Phys.* **17**, 338.
- KRÖNER, E. (1954) *Acta metall.* **2**, 301.
- KUHRT, F. (1952) *Z. Phys.* **131**, 185.
- LANGE, W. F. and AARONSON, H. I. (1979) *Metall. Trans.* **10A**, 1951.
- LEE, J. K. and AARONSON, H. I. (1975) *Acta metall.* **23**, 799, 809.
- LIN, J. (1968) *J. Chem. Phys.* **48**, 4128.
- LOTHE, J. and POUND, G. M. (1962) *J. Chem. Phys.* **36**, 2080; (1966) *Ibid.* **45**, 630; (1969) *Nucleation*, p. 109, Dekker, New York.
- NABARRO, F. R. N. (1940a) *Proc. Phys. Soc.* **52**, 90; (1940b) *Proc. R. Soc. A*, **175**, 519.
- NISHIOKA, K., SHAWYER, R., BIENENSTOCK, A., and POUND, G. M. (1971) *J. Chem. Phys.* **55**, 5082.
- PROBSTEIN, R. F. (1951) *J. Chem. Phys.* **19**, 619.
- REISS, H. (1952) *Ind. Eng. Chem.* **44**, 1284.
- REISS, H. and KATZ, J. L. (1967) *J. Chem. Phys.* **46**, 2496.
- REISS, H., KATZ, J. L., and COHEN, E. R. (1968) *J. Chem. Phys.* **48**, 5553.
- RODEBUSH, W. H. (1952) *Ind. Eng. Chem.* **44**, 1289.
- RUSSELL, K. C. (1968) *Acta metall.* **16**, 761; (1969) *Ibid.* **17**, 1123.
- TOLMAN, R. C. (1949) *J. Chem. Phys.* **17**, 333.
- TURNBULL, D. (1948) *Trans. Am. Inst. Min. (Metall.) Engr.* **175**, 774; (1950) *J. Chem. Phys.* **18**, 198.
- TURNBULL, D., and FISHER, J. C. (1949) *J. Chem. Phys.* **17**, 71.
- VOLMER, M. (1929) *Z. Elektrochem.* **35**, 555; (1939) *Kinetik der Phasenbildung*, Steinkopf, Dresden.
- VOLMER, M. and FLOOD, H. (1934) *Z. phys. Chem.* **170A**, 273.
- VOLMER, M. and WEBER, A. (1926) *Z. phys. Chem.* **119**, 227.
- ZELDOVICH, J. B. (1943) *Acta, Physicochim. URSS* **18**, 17.

## CHAPTER 11

# *Theory of Thermally Activated Growth*

### 53. GROWTH CONTROLLED BY PROCESSES AT THE INTERFACE

This chapter is concerned with the descriptive theory of growth in condensed phases. In Chapter 1, we classified transformations by considering the operative growth mechanism, and we shall now discuss the various types of thermally activated growth in more detail. Martensitic growth is treated separately in Part II, Chapter 21, and growth in liquid-solid transformations, which is usually controlled by heat flow, is discussed in Part II, Sections 66 and 67. The theory of the growth of a crystal from the vapour, which is more highly developed than are other examples of interface controlled growth, is also described separately in Part II, Chapter 13.

We begin by considering the type of growth which is controlled by processes in the immediate vicinity of the interface. For convenience we may suppose that there is a negligible change of volume and no change of composition as the boundary moves; this is nearly true for many transformations, and even in interface controlled precipitation processes, the boundary conditions of the diffusion equation may be such that the composition gradients can be virtually ignored.

There are two possible ways in which an incoherent crystal boundary might migrate normal to itself. In the first of these, atoms are able to cross the interface, adding themselves to the crystalline region on one side of the interface, simultaneously and independently at all points of the interface. This means that there is continuous growth at all points of the boundary. In the second mechanism, the interface is stepped on an atomic scale, and atoms are transferred from one phase to the other only at these steps. The interface then grows by the lateral motion of the steps, an element of surface undergoing no change until a step passes over it, when it moves forward through a distance equal to the step height.

In discussing the free surface of a crystal we have already pointed out that singular surfaces (surface orientations corresponding to sharp minima of surface free energy) will usually be atomically flat, whereas other surfaces will consist of atomic facets, or stepped sections of singular surfaces. Isolated atoms adding to a crystal on one side of a singular interface will be unstable and will tend to be removed again; the motion of such an interface thus requires a step mechanism, as we discuss in detail in Part II, Chapter 13. On

the other hand, it is usually considered that continuous growth of a non-singular interface is possible, provided the atomic disordering (faceting) is sufficiently extensive for atoms to be able to find a stable position in all parts of the interface.

The distinction between stepped and continuous growth is sometimes made on the basis of the diffuseness of the interface, i.e. the degree of atomic disorder and the extent of the transition region, rather than on the grounds of the singularity of the surface free energy. Of course, as we have pointed out in Chapter 5, these two criteria often go together, inasmuch as a singular interface is not disordered and is likely to be atomically sharp, whilst a non-singular interface is much more likely to be diffuse. However, Cahn (1960) has pointed out that although similar, the two distinctions are not necessarily equivalent, and he has derived a more general condition for deciding whether growth will be continuous or stepped. This criterion depends on the driving force (difference in free energy of the bulk phases on the two sides of the interface).

In introducing the idea of a step in Section 18, we assumed the interface to be sharp. However, a more general definition of a step may be given as the transition between two adjacent areas of surface which are parallel to each other and have identical atomic configurations, and which are displaced from each other by an integral number of lattice planes. This includes steps in diffuse interfaces, as already briefly mentioned for a special case on p. 294, even though the step height may be less than the thickness of the interface.

Cahn's theory of growth is based on the idea that a lateral growth (step) mechanism will be required whenever the interface is able to attain a metastable equilibrium configuration in the presence of the driving force. It will then tend to remain in such a configuration, advancing only by the passage of a step which does not change the configuration. If no such metastable equilibrium is possible, the boundary will move forwards continuously. This is thus a rather general thermodynamic criterion for growth mechanism, in contrast to the more atomistic theory based on the nature of the interface. The application of the driving force distinction, however, brings out the importance of the interface structure, which was previously recognized in a semi-intuitive way.

We shall not reproduce the details of Cahn's calculations here, but only quote the main results. For all types of interface, continuous growth is possible at a sufficiently large driving force, but stepped growth is required for driving forces less than some critical value. The magnitude of the critical driving force is very dependent on the nature of the interface; for very diffuse interfaces, it is so low that almost any driving force will lead to continuous growth, whereas for sharp interfaces it may be so high that it is never achieved in practice. Although the intuitive feeling that non-singular interfaces should always be able to grow continuously is not quite justified, the driving force required for such growth may be less than that attributable to the deviation of the actual crystal shape from the equilibrium (Wulff) shape. In other cases, it is possible that non-singular interfaces may grow by a step mechanism.

Let the position of the interface be given relative to some fixed lattice plane by a parameter  $z$ , and let the surface free energy of the interface as a function of  $z$  be

$$\sigma(z) = \{1 + \varphi(z)\}\sigma_0, \quad (53.1)$$

where  $\sigma_0$  is the minimum value of  $\sigma$ . Now let the interface move forward through a distance  $\delta z$ , so that the change in free energy per unit area of boundary is

$$\delta G = \left[ \left( (\Delta g^{\alpha\beta}/v) + \sigma_0 \frac{d\varphi}{dz} \right) \right] \delta z, \quad (53.2)$$

where  $-\Delta g^{\alpha\beta}$  is the driving force per atom and  $v$  is the atomic volume. The step growth mechanism is required only if for some value of  $z$  the change in free energy is zero. If the interface region contains a large number of lattice planes, Cahn shows that the maximum value of  $\sigma_0(d\varphi/dz)$  is  $\pi\sigma_0\varphi_{\max}/d$ , where  $\varphi_{\max}$  is the maximum value of  $\varphi$  and  $d$  is the interplanar spacing of the lattice parallel to the interface. If the negative of  $\Delta g^{\alpha\beta}/v$  is greater than this maximum variation in surface free energy, the interface should be able to advance continuously, since equilibrium configurations will not be attained. This gives a condition

$$\Delta g^{\beta\alpha}/v = -\Delta g^{\alpha\beta}/v > \pi\sigma_0\varphi_{\max}/d \quad (53.3)$$

for continuous growth. For diffuse interfaces,  $n$  lattice planes in extent, Cahn shows that the function  $\varphi(z)$  is given approximately by

$$\varphi(z) = (\pi^4 n^3/16) (1 - \cos 2\pi z/d) \exp(-\frac{1}{2}\pi^2 n) \quad (53.4)$$

and is very small when  $n$  is large. For very sharp interfaces, the maximum value of  $\varphi$  is of order unity, and this corresponds to very high critical driving forces according to the inequality (53.3).

Step growth will be discussed in detail in Part II, Chapter 13; for the remainder of this section we consider continuous growth on an atomic scale. A plausible model is one in which atoms cross the boundary independently of each other, each atom having to surmount an energy barrier in order to do this. In effect, we have already treated this problem in discussing the theory of nucleation in condensed assemblies (Section 49).

Consider first a phase transformation, in which a  $\beta$  region is growing into a metastable  $\alpha$  region. Figure 11.1 shows schematically the variation in free energy as an atom transfers from the  $\alpha$  phase to the  $\beta$  phase; the question of whether the activation energy  $\Delta_a g^*$  is to

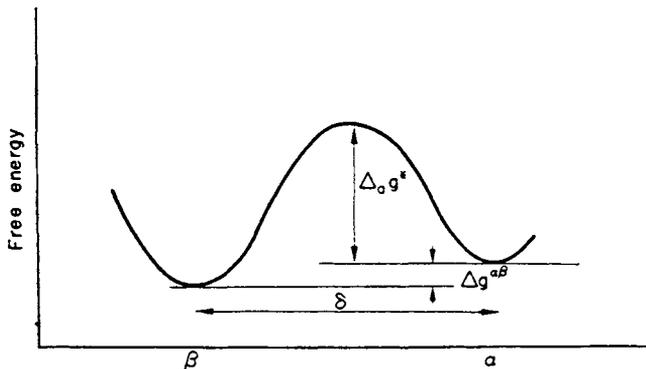


FIG. 11.1. Schematic variation of free energy for the activated transfer of atoms across the  $\alpha$ - $\beta$  interface.

be identified with the energy  $\Delta_a g^*$  of Section 49 is discussed below. In accordance with the theory developed in Chapter 3, the frequency with which an individual atom will transfer from the  $\alpha$  phase to the  $\beta$  phase is

$$\nu \exp(-\Delta_a g^*/kT),$$

where  $\nu$  is a characteristic frequency, and is given the value  $kT/h$  in Eyring's theory. As already explained, it is preferable to regard  $\nu$  as a frequency of known order of magnitude, the exact value of which is to be determined. This enables a straightforward physical interpretation to be given to both  $\Delta_a g^*$  and to  $\nu$ .

The frequency of the reverse transition of atoms from  $\beta$  to  $\alpha$  is similarly

$$\nu \exp\{-(\Delta_a g^* + \Delta g^{\beta\alpha})/kT\},$$

where, for simplicity, the frequency factors are assumed to be identical. The difference of these two expressions gives the net rate at which atoms transfer from the  $\alpha$  to the  $\beta$  phase. Thus if the distance across the interface is  $\delta^B$ , the velocity of the interface is

$$Y = \delta^B \nu \exp(-\Delta_a g^*/kT) [1 - \exp(-\Delta g^{\beta\alpha}/kT)]. \quad (53.5)$$

The formal similarity of this equation to the net rate of a chemical reaction (eqn. (15.11)) should be noted. At small values of the driving energy, we may expand the equation to give

$$Y \approx (\delta^B \nu/k) (\Delta g^{\beta\alpha}/T) \exp(-\Delta_a g^*/kT) \quad (53.6)$$

and the growth velocity is directly proportional to the difference in free energy of the two phases. The validity of this equation is limited to conditions where  $\Delta g^{\beta\alpha} \ll kT$ , a restriction which is usually satisfied in practice.

We have derived eqn. (53.6) by considering a specific model, but its form suggests that a much more general treatment in terms of the theory of Chapter 4 should be possible. Such a theory has been proposed by Machlin (1953), who pointed out that the motion of an interface separating two regions of different free energy may be regarded as a single irreversible process. The thermodynamical theory of irreversible processes can then be applied to the growth.

Consider a spherical element of the interface, having area  $O$  and radius of curvature  $r$ . If this moves normal to itself through a distance  $\delta r$ , the free energy change will be

$$\begin{aligned} \delta G &= (O \Delta g^{\alpha\beta}/v) \delta r + \sigma \delta O \\ &= -O \delta r \{ \Delta g^{\beta\alpha}/v - (2\sigma/r) \}, \end{aligned}$$

where  $v$  is the volume per atom and  $\sigma$  the surface energy of unit area of the interface. In accordance with the theory developed in Section 15, we require an expression for the rate of production of irreversible entropy as the boundary advances. This is given by  $-(1/T)(\partial G/\partial t)$ , so that the irreversible entropy per unit area of the boundary increases at a rate

$$\frac{d_i S}{dt} = \frac{1}{T} \left\{ \frac{\Delta g^{\beta\alpha}}{v} - \frac{2\sigma}{r} \right\} Y. \quad (53.7)$$

This equation has the expected form of a "force" and a conjugate "flux"  $\Upsilon$ , and the theory assumes that two such quantities are linearly related, so that

$$\Upsilon = (C_{13}/T) \{ \Delta g^{\beta\alpha}/v \} - (2\sigma/r). \quad (53.8)$$

For a plane interface, a suitable choice of  $C_{13}$  gives eqn. (53.6). For a curved interface we see that we should strictly include the change in surface energy in the overall free energy change used as the "driving force" for the motion. This is usually unimportant when  $r$  becomes large, but is always significant in grain growth, where it provides the only driving force.

The impression that the above theory gives a rather more general justification of the assumption that the boundary velocity is proportional to the free energy released when unit area of the boundary moves through unit distance, must be regarded as misleading. There is no necessity for the force and flux of eqn. (53.7) to be linearly related, and the validity of such a relation can only be settled experimentally. Eqn. (53.6) was derived on the basis of a particular atomic model, and will be correct if the real atomic processes approximate to those envisaged in the model. The corresponding equation (53.8) is justified only on the rather vague grounds that there are many circumstances in which the rate of entropy production is given by an equation like (53.7), and in which there is a proportionality between the "force" and the "flux". Quite evidently, there are other atomic models which do not give this linear relation between velocity and driving force. For example, Hillig and Turnbull (1956) considered as a very simple approximation that for step growth centred on dislocations (see Part II, Chapter 13) eqn. (53.6) should be multiplied by the fraction of boundary sites which are at growth steps. In a spiral step, the minimum radius of curvature cannot be smaller than the critical size of a two-dimensional nucleus, so that the total length of step in unit area is a function of the driving force. Thus in this model, the growth velocity is proportional to the square of the undercooling at the interface, whereas eqn. (53.8) predicts it is linearly proportional to the undercooling. A rather different example is provided by the diffusion controlled linear growth of dendrites (see Section 54). Some theories require a proportionality between growth velocity and driving force (or undercooling of the parent phase), but this assumption appears to be incorrect.

The quantity in curly brackets in eqn. (53.8) is the net change in free energy per unit volume produced by migration of the interface. If we write  $\Delta g^{\Upsilon}$  as the corresponding free energy change per atom, we may write the interface velocity as

$$\Upsilon = M^B \Delta g^{\Upsilon} \quad (53.9)$$

as the formal equivalent of either (53.6) or (53.8). We call  $M^B$  the mobility of the boundary. This concept of a boundary moving at a speed given by the product of a mobility and a driving force may be compared to the similar formal representation of a diffusion flow as the product of the mobility of an atom and the thermodynamic force acting on it (p. 384).

Returning to eqn. (53.6) we now have to consider the magnitude of  $\Delta_\alpha g^*$ . In general, this may be equal to the free energy of activation for the migration of atoms within the  $\alpha$  phase, but it can scarcely be greater than this, since the interface is disordered relative to the  $\alpha$  structure. If the boundary between the  $\alpha$  and  $\beta$  phases is incoherent, as it will usually be for

macroscopic crystals in a non-martensitic transformation,  $\Delta_a g^*$  may reasonably be expected to approximate to the activation energy for grain boundary diffusion, rather than lattice diffusion. Under these circumstances,  $\Delta_a g^*$  may be appreciably smaller than the activation free energy for growth at the nucleation stage  $\Delta_a g^*$ , since the latter will frequently relate to coherent or semi-coherent growth. In a dislocation-free region,  $\Delta_a g^*$  will then correspond to the energy for migration through the lattice.

For an incoherent boundary, the activation energy for growth should be independent of the relative orientations of the two crystals on each side of the boundary, or of the boundary itself, and the same should be true for a high angle grain boundary. Experiments on recrystallization and grain growth, however, show that some high angle grain boundaries of special orientations may be much more mobile than the other random boundaries. Very mobile boundaries appear to be "coincidence site boundaries" and the relative mobilities of special and random boundaries may be sensitive to very small quantities of impurities. These effects are discussed more fully in Part II, Chapter 19, but they show that the boundary motion can often not be represented as a simple thermally activated process, as assumed above. For boundary motion in grain growth, the driving force is almost independent of temperature, and the above theory implies that  $\ln \dot{\gamma}$  should be a linear function of  $1/T$ , the slope of which gives the heat of activation. Activation energies measured in this way are often much larger than is predicted by the above theory, and Mott (1948) suggested that atoms may be activated in groups rather than singly. The more probable explanation is that if impurities or inclusions are present, the temperature dependence of the growth rate does not give the activation enthalpy for boundary movement.

In recrystallization, the free energy  $\Delta g^{bz}$  will be replaced by the difference in free energies per atom in the strained and strain free regions. During grain growth there is no volume free energy term of this kind, and  $\dot{\gamma} = \partial r / \partial t$  is negative; the boundary migrates towards its centre of curvature. As the grain structure gradually coarsens, the effective driving force decreases, and hence the mean growth rate decreases as  $1/r$ . These ideal kinetics are usually observed only in zone-refined metals. It is to be noted that according to eqn. (53.8), interface controlled growth is strictly linear only when the interface is plane.

Returning to the application of eqn. (53.6) to phase transformations, we see that for a transition on heating  $\dot{\gamma}$  increases continually with increasing temperature. For a transition on cooling, the rate is zero at the transition temperature (eqn. (53.5)), increases to a maximum value, and then ultimately decreases again because of the overriding influence of the  $\exp(-\Delta_a g^*/kT)$  factor. This behaviour is observed in many transformations, e.g. that in pure tin, where the growth rate can be measured reasonably accurately. However, it is often difficult to obtain numerical agreement with theoretical predictions, and this has led to various suggestions for modifying the theory. One possibility is that the boundary mobility is restricted by inclusions which act as obstacles, or by impurities segregated to the boundary region, and this effect is known to be important for grain boundaries (see Section 83). Alternatively, atoms or molecules may be unable to transfer from one phase to another at all parts of the boundary, but only at certain active sites. These sites could be places where the local activation energy for crossing the boundary is reduced, perhaps because of the presence of imperfections in the boundary, or places where the local driving force is

increased. The latter possibility corresponds to the presence of steps in the interface, as discussed above in general terms, and this is the situation in growth from the vapour phase.

If the fraction of active sites is  $x_a$ , and the transfer of atoms is treated as before, the growth velocity of eqn. (53.6) will simply be multiplied by  $x_a$ . This has no effect on the functional form of the curves if  $x_a$  is independent of temperature, but it is more plausible to assume that the temperature dependence of  $x_a$  is important. Becker (1958) has given a theory of this kind for the growth of tin in which  $x_a$  decreases as the temperature increases.

When the variation of the driving force with temperature is known, it may be possible to deduce the heat of activation from experimental measurements. Writing

$$X = \Upsilon/\delta^B v [1 - \exp(-\Delta g^{\beta\alpha}/kT)]$$

and assuming that  $\Upsilon$  is equal to  $x_a$  times the right-hand side of (53.5), we have

$$\frac{d(\ln X)}{d(1/T)} = -\frac{\Delta_a h^*}{k} - \frac{T^2}{x_a} \frac{dx_a}{dT}. \quad (53.10)$$

If a graph of  $\ln X$  against  $1/T$  is a straight line, it follows that the second term on the right of (53.10) is either negligible in comparison with  $\Delta_a h^*/k$ , or is constant independent of temperature. The latter seems improbable, so that an experimental straight line plot may give the true heat of activation  $\Delta_a h^*$ . In view of the other possibilities mentioned above however, and of known examples from grain growth experiments in which the measured activation energy varies with purity, this assumption cannot be made with much confidence. The data required for this type of analysis are not available for many phase transformations.

#### 54. DIFFUSION-CONTROLLED GROWTH

In the preliminary discussion of isothermal transformation kinetics, given in Section 4, we assumed a linear growth law (constant growth rate). This assumption is appropriate under all circumstances in which the interface advances into a region of matrix of constant mean composition, so that the rate of the process controlling the growth is independent of the interface position, and hence of the time.

This situation does not necessarily apply when an isolated precipitate particle ( $\beta$ ) grows into a phase ( $\alpha'$ ) of different composition. If  $\beta$  is richer in solute atoms than the equilibrium  $\alpha$ , there may be a region depleted in solute formed around the  $\beta$  particle as growth proceeds; if the equilibrium  $\alpha$  contains more solute, then the region around the  $\beta$  particle may be enriched. In either case, the continual growth of the particle requires chemical diffusion in the surrounding  $\alpha'$ , and as the particle increases in size, the effective distances over which diffusion takes place may also increase. When the particle is first formed, it is probable that processes near the interface will control the net rate of growth, but the volume diffusion will eventually become the dominant factor if the interface mobility is reasonably high. The particle will then grow just as fast as the diffusion rate allows.

In contrast to the linear growth law obtained when processes near the interface are slowest, the position of the interface may be proportional to the square root of the time when diffusion is decisive. This may be seen from dimensional arguments, since (with certain

simplifying assumptions) the controlling diffusion equation and all the boundary conditions are homogeneous in the concentration. The concentrations of solute  $B$  atoms in the equilibrium  $\alpha$  and  $\beta$  phases  $c^\alpha$  and  $c^\beta$ , and the initial concentration in the metastable  $\alpha$  phase  $c^m$ , thus enter into the growth equations only in dimensionless combination. The linear dimensions of the  $\beta$  region are then functions only of the diffusion coefficient and the time, and possibly of each other. If all linear dimensions increase in the same way, each such dimension must be proportional to  $(Dt)^{1/2}$  and the growth is parabolic with a velocity varying as  $(D/t)^{1/2}$ . Alternatively, if some critical dimension  $r$  is constant, an increasing dimension may be proportional to  $Dt/r$  and the growth is linear with a constant velocity. In the first case, the effective diffusion distance is proportional to  $(Dt)^{1/2}$ , as in Section 40, whereas in the second case, this distance is constant and proportional to  $r$ .

In any transformation involving long-range transport, the diffusion equation

$$\frac{\partial c(\mathbf{x}, t)}{\partial t} = D \nabla^2 c(\mathbf{x}, t) \quad (54.1)$$

must be satisfied at each point in the metastable phase, the concentration  $c(\mathbf{x}, t)$  being a function of both position  $\mathbf{x}$  and time  $t$  (cf. eqn. (40.15)). The diffusion coefficient  $D$  will be assumed to be independent of concentration; the case of non-constant  $D$  is considered separately below. If there is no precipitate present at time  $t = 0$ , there is an initial condition

$$c(\mathbf{x}, 0) = c^m. \quad (54.2)$$

In order to find the distribution of solute from eqn. (54.1), we have to specify additional boundary conditions, and these are not immediately obvious. It seems reasonable to assume that there is a constant concentration of solute in the region of the  $\alpha'$  phase immediately adjacent to a growing particle. If the interface is mobile, this concentration cannot deviate much from the equilibrium  $\alpha$  concentration, and in the limit in which growth is entirely diffusion controlled, it may be put equal to this equilibrium concentration. However, if the interface has an appreciable curvature, the compositions of the  $\alpha$  and  $\beta$  phases in equilibrium are displaced from the values given by the phase diagram because of the Gibbs–Thomson effect. It is often convenient to make the simplifying assumption that  $c^\beta$  is independent of  $r$ , and we introduce the notation  $c_\infty^\alpha$  for the concentration of solute in equilibrium with  $c^\beta$  across a planar interface whilst  $c_r^\alpha$  will denote the local equilibrium composition which changes with the curvature. The distinction between  $c_\infty^\alpha$  and  $c_r^\alpha$  is important only for particles not much larger than the critical nucleus size, and is neglected in many growth theories. The assumption of local equilibrium is one limiting form of this boundary condition; the other limiting form is for a very sluggish interface, in which case the concentration near the boundary is almost equal to  $c^m$ . This leads to interface controlled growth. In any real transformation, the boundary condition must lie between these extremes, but we are often justified in approximating it by one or other of them.

A second boundary condition may be needed to ensure that the solute concentration behaves properly in regions between growing  $\beta$  particles. Finally, the fact that we have a moving boundary must be incorporated into the diffusion equation. This is done by finding

a relation between the velocity of the interface, and the flux of solute to or from the interface region.

When dealing with interface controlled linear growth, the theory of the growth rate is quite distinct from the theory of the mutual interference or impingement of growing regions. In any direction, the growth rate is constant until impingement occurs, after which it is zero. This is not so with diffusion limited growth. The interference of different growing crystals results from competition for the available excess solute, and becomes important when the diffusion fields of the two particles begin to overlap to an appreciable extent. This interference is thus part of the diffusion problem, and results in a growth rate which gradually becomes zero. The interference of different  $\beta$  regions is implicit in the second boundary condition, which deals with the concentration in  $\alpha'$  regions remote from the  $\beta$ . The two ways in which growing particles can interfere with one another are sometimes called "hard" and "soft" impingement respectively.

We find it convenient to discuss the formal kinetics of isothermal transformations separately in Chapter 12, and the effects of hard impingement are discussed there in a similar manner to the elementary treatment already given in Section 4. Although soft impingement is not really to be distinguished from the general problem of diffusion limited growth, we shall make an equivalent separation of this problem. The theory of diffusional growth in this section will thus be effectively confined to the growth of an isolated particle in an infinite matrix. Exact treatments of the diffusion problem for a sphere have been given by Zener (1949) and Frank (1950), and for particles of more general shape by Ham (1958, 1959) and Horvay and Cahn (1961). We shall begin with the earlier and much simpler treatment due to Zener, since Ham has shown that his method gives the correct result in most cases. All of these theories assume that the interfaces are smooth so that the diffusion flux is constant over a planar interface: the problem of diffusion-controlled growth of a stepped interface is considered later.

Zener considered one-dimensional growth (the thickening of a plate), isotropic two-dimensional growth (the radial growth of a cylinder) and three-dimensional growth (the growth of a sphere). His treatment assumes that the phases at the interface have equilibrium compositions, any variation with the curvature of the interface being neglected. The variation of composition along a line normal to the interface is assumed to be represented by Fig. 11.2.

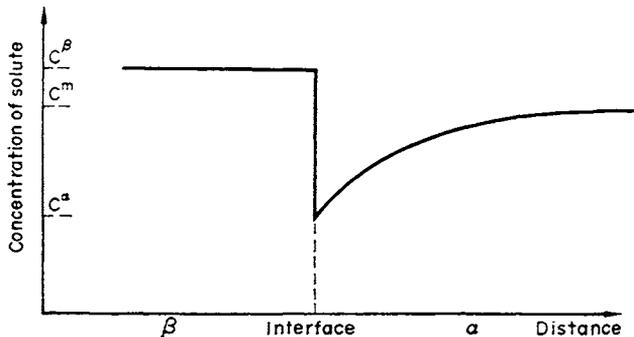


FIG. 11.2. Schematic variation of solute composition assumed in theory of diffusion controlled growth.

The precipitated  $\beta$  phase may be richer or poorer in  $B$  atoms than the  $\alpha$  phase, and the corresponding net diffusion current of  $B$  may be towards or away from the  $\beta$  particle. In either case, the concentration  $c^m$  is intermediate between  $c^\alpha$  and  $c^\beta$ . Changes in volume per atom as the  $\beta$  phase grows are ignored.

Let  $r$  be a coordinate normal to the interface, measuring the semi-thickness of the plate, or the radius of the cylinder or sphere. Then when the boundary is in a position  $r = r^t$ , the diffusion flux of  $B$  atoms across unit area in a time  $\delta t$  will be

$$D \left( \frac{\partial c}{\partial r} \right)_{r=r^t} \cdot \delta t,$$

where the concentration  $c$  of atoms in the  $\alpha$  phase is a function of  $r$  and  $t$ . If the interface advances a distance  $\delta r$  during this time, the extra number of solute atoms in a volume  $\delta r$  will be  $(c^\beta - c^\alpha) \delta r$ . These atoms can only come from the diffusion flow, so the two expressions may be equated. The limit of  $\delta r / \delta t$  gives the growth velocity, which is

$$\Upsilon = \frac{dr^t}{dt} = \frac{D}{c^\beta - c^\alpha} \left( \frac{\partial c}{\partial r} \right)_{r=r^t}. \quad (54.3)$$

It is sometimes convenient to represent the concentration gradient in this equation (rather loosely) as a ratio  $\Delta c / y^D$ , where  $\Delta c$  is the difference in solute concentration in matrix regions near to and remote from the interface, and  $y^D$  is an effective diffusion distance. The equation then becomes

$$\Upsilon = \frac{\Delta c}{c^\beta - c^\alpha} \frac{D}{y^D}. \quad (54.4)$$

We have already used the fact that eqns. (54.3) and (54.4) are homogeneous in the composition terms in our dimensional argument at the beginning of this section, and certain qualitative conclusions about the growth rate may be reached by consideration of (54.4). If the phase boundary is planar,  $y^D$  must continually increase as the boundary advances, leading to a diminishing growth rate, as mentioned above. For an interface which has a convex curvature towards the matrix,  $y^D$  will be approximately proportional to the radius of curvature. The increasing diffusion distance will thus also lead to a diminishing rate for the growth of a spherical particle or the radial growth of a cylinder. The growth conditions are not obvious, however, for the lengthening of a cylinder or the edgewise growth of a plate.

When a plate grows in its own plane, or a needle extends, it seems intuitively probable that the interface is always moving into fresh regions of constant composition, so that the growth rate is constant. This is certainly true if the rate is controlled by interface processes, composition gradients in the matrix being then virtually non-existent, but Zener (1949) and Wert (1949) implied that linear growth rates controlled by diffusion may also be obtained for particles of these shapes. A constant growth rate with diffusion control implies the existence of a steady-state solution to the diffusion equation, and from eqn. (54.4) we see that this should be possible if the growth conditions are such that the radius of curvature at the edge of a plate or the tip of a needle remains constant. Zener considered a flat plate to have such a constant radius equal to half the plate thickness, but he noted that the

restriction of the growth theory to dimensions much larger than those of a critical nucleus prevents us from applying eqn. (54.4) indiscriminately to a very thin plate, even though at first sight the reduction in  $y^D$  gives a high diffusion controlled rate of growth to such a plate. When the radius of curvature is small, the change in chemical free energy has to be subtracted from the chemical driving force, and there is a minimum radius below which growth is not possible. Another way of expressing this correction is that the effective difference in pressure across the curved interface changes the concentration of solute in the matrix with which the particle is in equilibrium (Gibbs–Thomson effect), and for the critical radius of curvature the particle is in equilibrium with the average composition of the matrix.

Steady-state solutions of the diffusion equation are important in the theory of discontinuous precipitation and eutectoidal decompositions (see Section 55), and also in such problems as the inward growth of a Widmanstätten plate from a grain boundary. More recent work has made it clear that there are two types of solution to be considered in relation to variations in particle shape. Plates and needles may be represented by prolate and oblate spheroids respectively, so that a continuous change of shape is obtained mathematically by varying the axial ratio of a family of ellipsoids with two equal axes. These particles all have closed form, and may thus be enclosed by a matrix. An alternative representation of shape variations is given by a family of paraboloids of elliptical cross-section, but this can only be used for dendrites, or for crystals growing in from a grain boundary. The limiting forms of the flat disc and the circular cylinder might apply to either enclosed or dendritic particles, but Zener's assumptions are incorrect for enclosed particles and rather doubtful for dendritic forms.

According to Zener and Wert's assumption, the eccentricity of an ellipsoidal particle increases during diffusion controlled growth. An opposite possibility was suggested by Doremus (1957) who assumed that a spherically symmetrical diffusion field will be established around any particle, independent of its shape. This implies that  $(\partial c/\partial r)$  in eqn. (54.3) is constant over the surface of a  $\beta$  particle, and growth in all directions is parabolic, the particles becoming more spherical as they grow. An exact treatment first given by Ham (1958) shows that both these predictions are incorrect; subject to certain assumptions, the particles have a fixed eccentricity and do not change shape.

We now have to obtain a more quantitative treatment of parabolic growth by evaluating eqn. (54.3). This means that  $c(r, t)$  has to be found by solving the general diffusion eqn. (54.1) which takes the form

$$\partial c/\partial t = D(\partial^2 c/\partial r^2) + (j-1)(D/r)(\partial c/\partial r), \quad (54.5)$$

where  $j = 1, 2,$  or  $3$  for one-, two-, or three-dimensional growth respectively. The boundary conditions for the solution of this equation are (54.2) and

$$c(r^j, t) = c^\alpha, \quad (54.6)$$

which corresponds to the concentration at the interface always being  $c^\alpha$ . Note that  $r^j$ , in this second condition, is itself a function of  $t$ .

Zener showed that if  $c^\alpha$  has the constant value  $c_\infty^\alpha$  the solution of eqns. (54.5) and (54.6) is

$$c = c^m + (c_\infty^\alpha - c^m) \frac{\varphi_j\{r/(Dt)^{1/2}\}}{\varphi_j\{r^j/(Dt)^{1/2}\}}, \tag{54.7}$$

where 
$$\varphi_j(x) = \int_x^\infty \xi^{1-j} \exp(-\xi^2/4) d\xi.$$

When  $j = 1$ , this reduces to a standard solution of the diffusion equation similar to those already given in Section 40. Since  $r^j \propto (Dt)^{1/2}$ , the quantity  $r^j/(Dt)^{1/2}$  is a dimensionless function of the concentrations which determines the growth velocity; we write it  $\alpha_j$  and relate it to the dimensionless supersaturation  $\bar{\alpha}$ , where

$$\bar{\alpha} = \frac{c^m - c_\infty^\alpha}{c^\beta - c_\infty^\alpha}. \tag{54.8}$$

From eqn. (54.3) we now find

$$\frac{1}{\bar{\alpha}} \frac{dr^j}{dt} = \frac{(\alpha_j)^{1-j}}{\varphi_j(\alpha_j)} \exp(-\alpha_j^2/4) \cdot D^{1/2} t^{1/2},$$

and hence

$$(\alpha_j)^j = \{2\bar{\alpha} \exp(-\alpha_j^2/4)\}/\varphi_j(\alpha_j). \tag{54.9}$$

This equation is an implicit expression for  $\alpha_j$  in terms of the quantity  $\bar{\alpha}$ , and for any supersaturation, the growth law is given by

$$r^j = \alpha_j (Dt)^{1/2}. \tag{54.10}$$

The relation in eqn. (54.10) is exact within the limitations of the assumptions mentioned above, but  $\alpha_j$  is not a simple function of  $\bar{\alpha}$ , varying from 0 to  $\infty$  as the latter varies from 0 to 1. Zener showed that simpler equations for  $\alpha_j$  could be obtained by using asymptotic expansions of  $\varphi_j(\alpha_j)$ , valid in the limits  $\alpha_j \ll 1$  and  $\alpha_j \gg 1$ . For one-dimensional growth, this gives

$$\left. \begin{aligned} \alpha_1 &= (2/\pi^{1/2})\bar{\alpha} & (\alpha_1 \ll 1), \\ \alpha_1 &= \{2(c^\beta - c_\infty^\alpha)/(c^\beta - c^m)\}^{1/2} & (\alpha_1 \gg 1), \end{aligned} \right\} \tag{54.11}$$

and for three-dimensional growth

$$\left. \begin{aligned} \alpha_3 &= (2\bar{\alpha})^{1/2} & (\alpha_3 \ll 1), \\ \alpha_3 &= \{6(c^\beta - c_\infty^\alpha)/(c^\beta - c^m)\}^{1/2} & (\alpha_3 \gg 1). \end{aligned} \right\} \tag{54.12}$$

Zener also showed that it is convenient to express  $\alpha_j$  in terms of different (“natural”) concentration parameters, which may be obtained by considering approximate solutions. For one-dimensional growth, such a solution follows from the assumption that the concentration decreases linearly from  $r = r^j$  to some higher value of  $r$ , after which it is constant at  $c = c^m$ . The condition that the total extra number of solute atoms in the  $\beta$  region has

come from the surrounding depleted region then fixes the outer limit of the latter, and hence the assumed constant value of  $\partial c/\partial r$ . This gives

$$\partial c/\partial r = (c^m - c_\infty^\alpha)^2/2(c^\beta - c^m)r^I$$

and on substituting into (54.3), we find

$$r^I \approx \frac{(c^m - c_\infty^\alpha)}{(c^\beta - c_\infty^\alpha)^{1/2}(c^\beta - c^m)^{1/2}} (Dt)^{1/2} = \alpha_1^* (Dt)^{1/2}. \quad (54.13)$$

The quantity  $\alpha_1^*$  is thus a natural parameter in which to express the growth law for a plane interface. The true growth parameter,  $\alpha_1$ , given by (54.9), is related to  $\alpha_1^*$  by

$$\alpha_1 = C\alpha_1^*,$$

where  $C$  varies from  $2/\pi^{1/2} \approx 1.13$  to  $\sqrt{2}$  as  $c^m$  varies from  $c_\infty^\alpha$  to  $c^\beta$  (compare asymptotic expressions (54.11) above). Equation (54.9) was solved graphically by Zener, and plotted in the form of a smooth  $\alpha_1 - \alpha_1^*$  relation, but it is clear that eqn. (54.13) is a sufficient approximation for all practical purposes. For the growth of a spherical particle, an estimate of  $(\partial c/\partial r)$  may be made in the same way. This is equivalent to supposing that the solute atoms are taken from a thin spherical shell surrounding the precipitate particle, and the concentration gradient is assumed constant across this shell. This will only be a good approximation when there is a high degree of supersaturation, so that

$$c^\beta - c^m \ll c^\beta - c_\infty^\alpha.$$

The result may be expressed in the form of a growth coefficient which is  $\sqrt{(3)}\alpha_1^*$ .

At low degrees of supersaturation, i.e. when

$$c^\beta - c^m \gg c^m - c_\infty^\alpha \quad (54.14)$$

the approximation of a linear change of concentration will not be valid. In this case, the limiting form of the steady-state solution for a stationary boundary is appropriate, so that the diffusion problem reduces to finding the solution of Laplace's equation

$$\nabla^2 c = 0 \quad (54.15)$$

obtained from (40.15) or (54.1) by setting  $\partial c/\partial t = 0$ . With the above boundary conditions, the solution for spherical growth is

$$c = c^m(1 - r^I/r) + c_\infty^\alpha r^I/r \quad (54.16)$$

so that

$$(\partial c/\partial r)_{r=r^I} = (c^m - c_\infty^\alpha)/r^I$$

and the effective diffusion distance (eqn. 54.4) is given by

$$y^D = r^I. \quad (54.17)$$

This gives a growth coefficient

$$\alpha_3^* = \sqrt{(2)}(c^m - c_\infty^\alpha)^{1/2}/(c^\beta - c^m)^{1/2} \quad (54.18)$$

so that there is an apparent difference between the "natural" growth parameters for small and large supersaturations. However at high supersaturations, the coefficient  $\sqrt{3}\alpha_1^* \simeq (\frac{3}{2})^{1/2}\alpha_3^*$  so that for both high and low supersaturations, the quantity  $\alpha_3^*$  forms a suitable parameter for the growth law. Zener's graphical computation shows that the true growth coefficient  $\alpha_3 = C\alpha_3^*$ , where  $C$  varies slowly from 1 to  $\sqrt{3}$  as  $c^m$  changes from  $c_\infty^a$  to  $c^\beta$  (see eqns. (54.12)).

When the approximation of the steady-state solution is valid, it may also be used to deduce the growth rate for some other geometries. An example is the thickening of a circular cylinder, where, however, it is necessary to replace the boundary condition at infinity by

$$c = c^m \quad \text{for} \quad r \geq r_e,$$

where  $r_e$  is an outer cut-off radius, large in comparison with the radius of the cylinder  $r^l$ . The solution of the steady state diffusion equation is then

$$c = c_\infty^a + (c^m - c_\infty^a) \ln(r/r^l) / \ln(r_e/r^l), \quad (54.19)$$

and the effective diffusion distance in eqn. (54.4) becomes

$$y^D = r^l \ln(r_e/r^l). \quad (54.20)$$

Solutions are also available for more complex geometries, such as a parabolic cylinder or a paraboloid of revolution (Trivedi and Pound, 1967).

As noted in Chapter 9, the diffusion coefficient is often a function  $D(c)$  of composition, and we consider next the effect of this variation on the computed growth rate. This problem has been considered by Trivedi and Pound (1967) and Atkinson (1967, 1968). The conservation equation (54.3) still applies with  $D$  having the value  $D(c_\infty^a)$  appropriate to the interface, but the diffusion equation has to be solved in the general form (40.15) rather than in the form (54.1).

Trivedi and Pound considered first the case in which the steady-state solution for a stationary boundary is an adequate approximation, i.e. when eqn. (54.14) is valid. The diffusion equation then reduces to the time-independent form

$$\nabla(D \nabla c) = 0 \quad (54.21)$$

and on making the substitution

$$\psi(c) = \{D(0)\}^{-1} \int_0^c D \, dc. \quad (54.22)$$

this becomes

$$\nabla^2 \psi = 0. \quad (54.23)$$

Also from (54.3)

$$\Upsilon = \frac{D(0)}{(c^\beta - c_\infty^a)} \left( \frac{\partial \psi}{\partial r} \right)_{r=r^l}, \quad (54.24)$$

which can be written in a form similar to (54.4) as

$$\Upsilon = \frac{D(0)}{(c^\beta - c_\infty^a)} \frac{\psi(c^m) - \psi(c_\infty^a)}{y^D}. \quad (54.25)$$

Since (54.23) has the same form as (54.15), the value of  $y^D$  appropriate to (54.25) is identical with that in (54.4) for the case of  $D$  constant, e.g.  $y^D = r^I$  for spherical growth. Moreover, it follows from (54.22) that (54.25) can be placed in the form (54.4) if  $D$  in the latter equation is given the value

$$\bar{D} = \int_{c_\infty^\alpha}^{c^m} D dc / (c^m - c_\infty^\alpha). \quad (54.26)$$

Thus the important result of this theory is that for a concentration dependent diffusion coefficient, the growth rate corresponds to that calculated for a constant diffusion coefficient evaluated as a weighted average over the composition range found in the matrix. This corrects earlier suggestions that either the value of the diffusion coefficient at  $c = c_\infty^\alpha$  or the maximum value in the range  $c_\infty^\alpha - c^m$  should be used in place of the constant  $D$  of the simple solutions.

Trivedi and Pound also considered the effect of variations in  $D$  with composition on the edgewise growth of a plate-shaped particle. The growth theory is more complex (see below), but with certain assumptions they were able to obtain numerical solutions for the growth rate. The computed values differed only slightly from those calculated from an assumed constant  $D$  given by the weighted average of (54.26).

Atkinson considered one-dimensional growth when the steady-state approximation is invalid. He assumed that the growth rate continues to be given by an equation of form (54.10), and he was able to obtain analytical solutions for the growth coefficient  $\alpha_1$  for a class of exact solutions originally considered by Philip (1960). An approximation to these solutions is given by Zener's theory with  $D$  replaced by an average value which is determined mainly by  $D(c_\infty^\alpha)$  and  $D(c^m)$ . In view of the difficulty of matching experimental results for  $D(c)$  with a form giving an exact solution, Atkinson (1968) also developed an iterative procedure for the numerical solution of the diffusion equation for one-dimensional growth. The one-dimensional form of eqn. (54.5) is first transformed into a moving system of coordinates with the origin fixed in the interface, and the Boltzmann substitution  $\lambda = x/t^{1/2}$  is then used to give a relation between  $c$  and  $\lambda$  which involves the experimental function  $D(c)/D(c_\infty^\alpha)$ . The differential equation is then solved numerically by a technique which involves beginning with an arbitrary value of the growth coefficient  $\alpha_1$  which defines the interface position through

$$r^I = \alpha_1 \{D(c_\infty^\alpha)t\}^{1/2} \quad (54.27)$$

(compare eqn. (54.10)) and iterating until a constant value is obtained for  $\alpha_1$ . However, as already indicated, it is probably adequate in most cases to use the solution for a constant diffusion coefficient with an appropriate average value of  $D$ .

We now turn to a brief consideration of the treatment by Ham. He uses a more complex mathematical procedure, in which the solute density is expanded in terms of the eigenfunctions of an appropriate boundary value problem. This enables him to obtain solutions of the diffusion equation for particles of any shape provided they are distributed periodically in the matrix. The form of the solution shows that it may be applied also to a non-uniform distribution, so that the method includes a proper treatment of the problem of soft impinge-

ment. We shall not describe the calculations in detail, but we shall summarize some of the important results; others will be mentioned in Chapter 12.

Ham assumes that the particles grow from initially negligible dimensions, that the concentration of solute in the matrix adjacent to the particle is constant at all parts of the surface, and that accumulation of solute occurs at the point of contact with the particle. As already mentioned, this leads to the result that particles growing as spheroids of any shape maintain constant eccentricity under diffusion limiting conditions. Each dimension of a particle is proportional to the square root of the time, and the volume varies as  $t^{3/2}$ , just as for a sphere. This contrasts with the earlier incorrect suggestions that the volume varies as  $t^{5/2}$  for plates and  $t^2$  for rods. Ham also showed that these results for spheroids may be obtained by Zener's approximate method. As described above, this involves the calculation of the concentration gradient by the steady-state solution of the diffusion equation near a particle of fixed dimensions equal to the instantaneous dimensions of the growing particle.

Ham also considered the problem of small rods and plates of finite initial dimensions which remain highly eccentric and do not alter their long dimensions appreciably during the diffusion growth. (This is appropriate, for example, if the shape is established by rapid interface controlled growth, as suggested above.) In both cases the volume increases as  $t^4$  during the subsequent diffusion growth. This contradicts the assumption of Zener and Wert that a plate thickens as  $t^{1/2}$ , as predicted from the above theory of one-dimensional growth. The theory is correct, but is not applicable to the thickening of a plate of finite initial dimensions since the actual diffusion field is then quite different from that of one-dimensional diffusion. There is an initial transient whilst the diffusion field is established, during which the growth changes from a  $t^{1/2}$  to a  $t^1$  law.

The existence of shape-preserving solutions does not, of course, imply that growth is necessarily of this form for all enclosed particles. If the interface processes are dependent on the orientation of the boundary, for example, or if thermal gradients exist in the matrix, the assumption of constant concentration at the boundary would not be valid. This would lead to a change of shape as the particle grew. Similar effects occur if the curvature of the boundary is large, since this displaces the equilibrium composition because of the Gibbs-Thomson effect. However the most important reservation, as Ham himself recognized, relates to the stability of a growth front under diffusion-limited conditions.

The problem of interface stability first came into prominence in work on the solidification of liquids, and in Part II, Chapter 14, we shall find that an initially planar solid-liquid interface may become non-planar under certain conditions of solute diffusion. Rather more extreme conditions then lead to "projections" on the solid becoming unstable and growing out rapidly into the liquid (dendritic growth). Comparable situations should arise in diffusion-limited growth in the solid state. The first systematic treatment of the problem of morphological stability was given by Mullins and Sekerka (1963, 1964), and this branch of growth theory has been important ever since their work.

We give here a necessarily over-simplified account of the theory of stability which is adapted primarily to the solid-state diffusion problem; the liquid-solid problem is similar but is complicated by heat flow and will be considered separately in Part II, Chapter 14.

Consider a planar interface  $x_1 = 0$ , moving with velocity  $\Upsilon$ , and superimpose on this interface a shape perturbation

$$x_1 = \delta(t) \sin \omega x_2, \quad (54.28)$$

where  $x_2$  defines a direction in the interface. The wavelength of this perturbation  $2\pi/\omega$  is not necessarily large compared with the amplitude  $\delta$ , so that the effect of curvature on the equilibrium composition can not be neglected. Figure 11.3 shows two extreme conditions; in (a) the composition in the  $\alpha$  phase remains constant along the perturbed interface, so that the concentration contours in front of the interface are also sinusoidal with diminishing amplitude, whilst in (b) the concentration contours remain planar, so that the composition varies along the non-planar interface. In Fig. 11.3(a) the concentration gradient in the  $x_1$  direc-

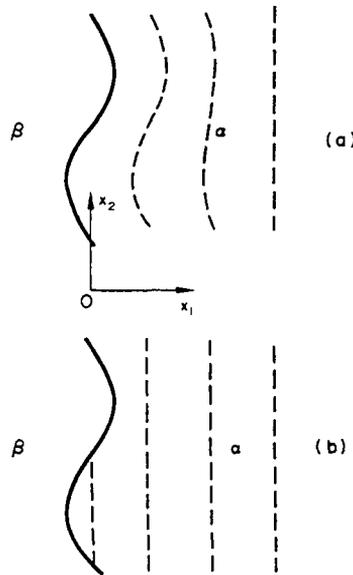


FIG. 11.3. To illustrate the instability of a planar interface when growth is diffusion-controlled (after Shewmon, 1965).

tion is increased ahead of the “hills” (i.e. the iso-concentration surfaces are compressed together) and reduced in front of the valleys, whilst in Fig. 11.3(b) the concentration gradient normal to the overall interface is constant.

If the situation is as envisaged in Fig. 11.3(a) the hills will tend to grow faster and the valleys more slowly compared with the growth rate of a flat interface. Thus the amplitude of the sine wave will tend to increase with time, and the flat interface is inherently unstable. However, as the amplitude increases so also does the curvature and the equilibrium composition of the  $\alpha$  phase begins to increase ahead of the hills and to decrease ahead of the valleys. Thus the Gibbs–Thomson effect exerts a stabilizing influence and tends to diminish the perturbation; the situation of Fig. 11.3(b) will be obtained, in principle, if the variation of composition along the interface is just that required for equilibrium of  $\alpha$  and  $\beta$  planes over a curved interface.

For small values of  $\delta$  the curvature is approximately  $d^2x_1/dx_2^2$ , so that the composition of the  $\alpha$  phase in equilibrium with the  $\beta$  becomes (see eqn. (22.38))

$$c^\alpha = c_\infty^\alpha [1 + \Gamma \delta(t) \omega^2 \sin \omega x_2]. \quad (54.29)$$

The composition variation in the  $\alpha$  phase may be obtained in the steady-state approximation (i.e. when (54.14) is valid) by writing down the general solution to Laplace's equation and imposing the boundary conditions (54.29). This gives (Shewmon, 1965)

$$c(x_1, x_2) = c_\infty^\alpha + [c_\infty^\alpha \Gamma \omega^2 - (\partial c / \partial x_1)_I] \delta \sin(\omega x_2) \exp(-\omega x_1) + x_1 (\partial c / \partial x_1)_I, \quad (54.30)$$

where  $(\partial c / \partial x_1)_I$  is the composition gradient at the interface. The velocity of the interface eqn. (54.3) is thus

$$V(x_2) = \frac{D}{c^\beta - c^\alpha} \left[ \left( \frac{\partial c}{\partial x_1} \right)_I - \left\{ c_\infty^\alpha \Gamma \omega^2 - \left( \frac{\partial c}{\partial x_1} \right)_I \right\} \omega \delta \sin \omega x_1 \right]. \quad (54.31)$$

The growth rate of a fluctuation, i.e. the velocity of a peak relative to that of the mean interface, is

$$\delta = \frac{D\omega}{c^\beta - c^\alpha} \left[ \left( \frac{\partial c}{\partial x_1} \right)_I - c_\infty^\alpha \Gamma \omega^2 \right] \delta. \quad (54.32)$$

According to this simple theory, the expression for the growth of a fluctuation contains a positive term, linear in  $\omega$  and  $(\partial c / \partial x_1)_I$ , and a stabilizing negative term which depends on  $\Gamma \omega^3$ . The critical wave number  $\omega_c$  above which  $\delta < 0$  and below which  $\delta > 0$  is given by

$$\omega_c = \{ (\partial c / \partial x_1)_I / c_\infty^\alpha \Gamma \}^{1/2} \quad (54.33)$$

and the maximum growth rate occurs at

$$\omega_1 = \omega_c / 3^{1/2}. \quad (54.34)$$

The growth rate is negative at large  $\omega$  (small wavelengths) because of the Gibbs-Thomson effect and is small at small  $\omega$  (large wavelengths) because of the large diffusion distance.

In their original paper, Mullins and Sekerka (1963) treated the growth of a sphere with a superimposed perturbation  $\delta_i Y_{lm}$  in the form of a spherical harmonic  $Y_{lm}(\theta, \varphi)$  of amplitude  $\delta_i$ . For a particular harmonic, they found the growth rate of the fluctuation to be given by

$$\dot{\delta}_i = \frac{D(l-1)}{(c^\beta - c^\alpha) r^l} \left[ \left( \frac{\partial c}{\partial r} \right)_{r_i} - \frac{c_\infty^\alpha \Gamma}{(r^l)^2} (l+1)(l+2) \right] \delta_i, \quad (54.35)$$

which is of the same form as (54.32). It follows from this equation that a harmonic of given  $l$  will grow if the radius of the sphere exceeds a value  $r_1(l)$  given by

$$r_1(l) = \left( \frac{1}{2}(l+1)(l+2) + 1 \right) r_c, \quad (54.36)$$

where

$$r_c = 2c_\infty^\alpha \Gamma / (c^m - c_\infty^\alpha) \quad (54.37)$$

is the critical radius at which a  $\beta$  precipitate is in equilibrium with the matrix of composition  $c^m$  (see eqn. (22.33)) and thus represents the critical nucleus size for this degree of super-

saturation (see Part II, Chapter 18). The first harmonic which distorts the spherical shape has  $l = 2$  and hence  $r_1 = 7r_c$ . A sphere perturbed by  $\delta Y_{20}$  is to first order an ellipsoid of eccentricity  $2\delta/r$ , and application of eqn. (54.35) to this shape when  $r \gg r_c$  (Gibbs–Thomson effect negligible) shows that the solution is shape preserving, i.e. a sphere perturbed only by this harmonic does not change shape as it grows. This result is thus consistent with Ham's solutions, which were obtained by neglecting the effect of curvature. More severe perturbations of the sphere give non-shape preserving solutions and for  $l = 3$ , the condition is

$$r_1 > 21r_c. \quad (54.38)$$

These solutions indicate that for both spherical and non-spherical particles instability is favoured by high supersaturations and large particle sizes. An individual sphere growing in a supersaturated matrix tends to dissolve for  $r < r_c$  and the spherical shape is unstable for  $r > r_1$ ; similar considerations apply to an ellipsoidal particle. Thus there should be a comparatively small range of sizes for which Ham's shape preserving solutions are valid; near  $r = r_c$ , the neglect of Gibbs–Thomson effects is a serious limitation and for  $r > r_1$  the shape preserving solution is unstable against fluctuations.

Shewmon (1965) pointed out that the experimental evidence suggests that stable polyhedral growth forms are found to persist over a much wider range of sizes and supersaturations than is implied by the above theory, and hence that additional stabilizing factors must be considered. Neglected effects include diffusion in the  $\beta$  phase if its composition is non-uniform, faster diffusion along the  $\alpha$ – $\beta$  interface than in the bulk  $\alpha$  phase, the effects of transformation stresses, and the possible existence of a slow interfacial reaction (i.e. growth which is not completely diffusion controlled); impurities, however, will increase the tendency to growth instability. Surface diffusion adds a negative term in  $\omega^4$  to eqn. (54.32), and this has the effect of decreasing  $\omega_c$  and thus extending the range of  $\omega$  over which a planar growth front is stable against fluctuations, but it does not remove the instability at small  $\omega$ . Shewmon concluded that transformation stresses have a rather weak stabilizing effect, and that observed non-dendritic growth forms in solid-state precipitation reactions probably indicate some degree of interface control of the growth rate.

A growth morphology which bears some resemblance to dendritic growth is the formation of long, thin needles or plates (Widmanstätten side plates) by inward growth from a grain boundary. Experimental observations show that the growth of such plates is linear rather than parabolic, so that it is necessary to look for steady-state solutions of the diffusion equation. One model of this type, due to Hillert (1957), is a development of Zener's theory for the growth of a plate of constant thickness or a needle of constant diameter; other more

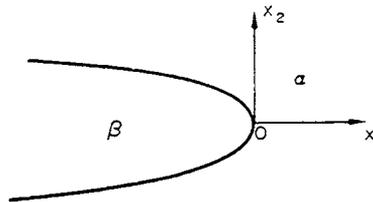


FIG. 11.4. To illustrate the Zener–Hillert growth model.

recent models consider plates of parabolic section. The Zener–Hillert model, which is shown in Fig. 11.4, is admittedly inexact, but it has assumed a prominent place in the literature and will therefore be discussed first.

Zener assumed that the tip of a needle is hemispherical and the edge of a plate is hemicylindrical so that the composition of the  $\alpha$  phase at the interface,  $c_r^\alpha$  is constant over the curved surface. In eqn. (54.4)  $\Delta c$  is now given by  $(c^m - c_r^\alpha)$  and it follows from eqns. (22.33), (22.38), and (54.37) that this is equal to  $(c^m - c_\infty^\alpha)(1 - r_c/r)$ . The composition term in the denominator of (54.4) is  $c_r^\beta - c_r^\alpha$  but it follows from the discussion of the Gibbs–Thomson effect in Section 22 that this is nearly equal to  $c_\infty^\beta - c_\infty^\alpha$ . The diffusion distance  $y^D$  is assumed to be proportional to  $r$  and hence to remain constant with time for a needle or plate of given thickness. The growth rate may thus be written

$$\dot{Y} = (D\bar{\alpha}/C_{14}r) \{1 - (r_c/r)\}, \tag{54.39}$$

where  $y^D = C_{14}r$  and  $\bar{\alpha}$  is given by (54.8).

The growth rate given by (54.39) is zero for  $r = r_c$  and tends to zero as  $r \rightarrow \infty$ . The maximum growth rate occurs when

$$r = 2r_c \tag{54.40}$$

and is given by

$$\dot{Y} = D\bar{\alpha}/4C_{14}r_c. \tag{54.41}$$

Note that eqns. (54.39)–(54.41) are valid whatever approximations are made to fix  $\Gamma$  in (22.33) or (22.38).

Equation (54.39) gives a relation between velocity and tip radius but does not predict either uniquely. Zener assumed that the observed velocity will be the maximum possible velocity, given by eqn. (54.41), and correspondingly that the diameter of a needle or the thickness of a plate will be given by  $2r = 4r_c$  from (54.40). This assumption is one of the more contentious points of the theory. In view of the approximate nature of the treatment, it is not possible to give a reliable estimate for  $C_{14}$ , but if eqn. (54.17) remains valid,  $C_{14} \sim 1$ .

In the theory of heat flow it is often convenient to use a dimensionless velocity parameter known as the Péclet number. The corresponding Péclet number  $\bar{p}$  for diffusional growth is obtained by replacing the thermal diffusivity by the diffusion coefficient and is defined as

$$\bar{p} = \dot{Y}r/2D \tag{54.42}$$

The Zener growth theory is given by the condition

$$\bar{\alpha}_{\text{eff}} = \bar{\alpha}\{1 - (r_c/r)\} = 2C_{14}\bar{p} \tag{54.43}$$

and the maximum velocity condition is

$$\bar{\alpha} = 4C_{14}\bar{p}, \quad r = 2r_c. \tag{54.44}$$

Hillert (1957) attempted to give a more rigorous treatment of the diffusion problem. In Fig. 11.4 composition gradients along  $x_3$  may be neglected in first approximation. Consider an origin in the moving boundary and let  $x_1, x_2$  be coordinates in the direction of growth and normal to the plane of the plate respectively. Then at any point, the rate of change of

concentration caused by diffusion processes is  $D\{\partial^2 c/\partial x_1^2 + (\partial^2 c/\partial x_2^2)\}$  (see eqn. (54.1)). The corresponding change of concentration due to the boundary velocity,  $\Upsilon$ , is  $\Upsilon(\partial c/\partial x_1)$ . If the growth of the plate is a steady-state process with constant  $\Upsilon$ , the concentration of any point with fixed coordinates relative to the moving edge of the plate remains unchanged. Thus we have the diffusion equation

$$D\{(\partial^2 c/\partial x_1^2) + (\partial^2 c/\partial x_2^2)\} + \Upsilon(\partial c/\partial x_1) = 0. \quad (54.45)$$

This may be contrasted with eqn. (54.5) which describes a growth process in which a steady state is not attained.

A solution of this equation is obtained by separating the variables, that is by making the assumption that

$$c(x_1, x_2) = g(x_1) h(x_2). \quad (54.46)$$

Note that  $c$  is not a function of the time in a steady-state solution, since the time has been eliminated by using a moving coordinate system. Substituting (54.46) into (54.45) gives

$$\frac{1}{g} \left\{ \frac{d^2 g}{dx_1^2} + \frac{\Upsilon}{D} \frac{dg}{dx_1} \right\} = -\frac{1}{h} \frac{d^2 h}{dx_2^2} = b^2$$

and the particular solutions are

$$g = \exp(-\lambda x_1), \quad h = \cos(b x_2), \quad (54.47)$$

where

$$\lambda = (\Upsilon/2D) (1 + \sqrt{[1 + 4b^2 D^2/\Upsilon^2]}) \quad (54.48)$$

and  $b$  can take any constant value. The complete solution is thus

$$c - c^\alpha = \int_0^\infty A(b) \exp(-\lambda x_1) \cos(b x_2) db. \quad (54.49)$$

The problem now reduces to the determination of the coefficients  $A(b)$  from the boundary conditions of the model, the assumption of local equilibrium at the interface, etc. Unfortunately this cannot be done because the boundary conditions are in fact not compatible with the existence of the steady-state solution. For this reason, the value of the whole calculation is rather doubtful, and it seems unlikely that the model is a good approximation to the real physical situation. Hillert avoids the difficulty by specifying an internal condition which satisfies the boundary conditions only in the  $x_2$  direction. The concentration in the plane  $x_1 = 0$  is assumed to fit the equation

$$c - c^\alpha = C_{15} \exp(-\pi x_2^2/16m^2). \quad (54.50)$$

where  $C_{15}$  and  $m$  are arbitrary constants. Although this is a plausible form for the concentration change, no real justification can be given for this assumption. If it is used to replace the real boundary conditions, however, a unique solution to the diffusion problem is obtained, the steady-state growth rate being given by

$$\Upsilon = \frac{D\bar{\alpha}}{2} \frac{(c^\beta - c_r^\alpha)}{(c^\beta - c^m)} \frac{1}{m} \left( 1 - \frac{m_c}{m} \right), \quad (m \simeq r). \quad (54.51)$$

This corresponds to Zener's equation (54.39) with

$$C_{14} = 2(c^\beta - c^m)/(c^\beta - c_r^\alpha).$$

After substituting into (54.43) and rearranging,

$$\bar{\alpha}_{\text{eff}} = 4\bar{p}(1 + 4\bar{p})^{-1}. \quad (54.52)$$

Trivedi and Pound (1969) refer to this as the Zener-Hillert equation and point out that at low supersaturations (small  $\bar{p}$ )

$$\bar{\alpha}_{\text{eff}} = 4\bar{p}, \quad (\bar{\alpha}_{\text{eff}} \rightarrow 0), \quad (54.53)$$

whereas at high supersaturations (large  $\bar{p}$ )

$$\bar{\alpha}_{\text{eff}} = 1 - (1/4\bar{p}) \quad (\bar{\alpha}_{\text{eff}} \rightarrow 1). \quad (54.54)$$

The Zener theory gives a linear relation between velocity and supersaturation, but this is true for Hillert's result only in the limit of (54.53). At very high supersaturations the Péclet number tends to infinity as  $\bar{\alpha} \rightarrow 1$ ; this is expected physically because as  $c^m \rightarrow c^\beta$ , the diffusion required becomes negligible. Previous treatments of the Hillert theory have generally approximated  $(c^\beta - c_r^\alpha)/(c^\beta - c^m)$  by unity and have thus concluded that it is equivalent to the Zener theory with  $C_{14} = 2$ . The maximum growth rate then occurs at  $r = 2r_c$  and  $\bar{\alpha}_{\text{eff}} = \frac{1}{2}\bar{\alpha}$ , but these results are valid only in the low supersaturation limit.

Hillert also gave a similar treatment for the growth of a needle-shaped particle and for the growth of a lamellar aggregate (pearlite). The pearlite problem, to which the steady-state solution is appropriate, is discussed in Section 55. The boundary conditions do not permit a steady-state solution for the growth of a cylindrical needle of constant thickness, and a similar approximate method has to be used. The maximum growth rate obtained by Hillert for a needle is 1.5 times that for a plate.

The minimum radius of curvature is given by the condition that the net free energy change is zero when a plate of this thickness extends. Let  $-\Delta g^f$  be the chemical driving force per atom for the growth of the plate. The chemical free energy released when unit length of the plate moves through a distance  $dx$  is thus  $-(y \Delta g^f/v) dx$  where  $y$  is the thickness of the plate and  $v$  is the volume per atom. The change in surface energy is  $2\sigma dx$ , and in the limiting case when this equals the chemical free energy released,  $y = 2r_c$ . Hence

$$r_c = -\sigma v/\Delta g^f. \quad (54.55)$$

An exact solution of the diffusion problem for a particle growing with constant velocity in one direction may be obtained by neglecting the Gibbs-Thomson effect of the curvature, which is emphasized in the Zener-Hillert model, and returning to the boundary conditions used by Ham. For the problem of dendritic growth controlled by heat flow, Papapetrou (1935) first pointed out the existence of such a linear solution when the dendrite has the shape of a paraboloid of revolution. The most complete treatment of this problem (Horvay and Cahn, 1961) is based on an analysis due to Ivantsov (1947) and shows that exact solutions can be obtained for a family of elliptical paraboloids, including as special cases needles,

or dendrites with aspect ratio of the ellipse near unity, and dendritic plates. For the case in which the shape is a parabolic cylinder, i.e. an infinite plate, the Ivantsov theory gives

$$\bar{\alpha} = (\pi\bar{p})^{1/2} \exp(\bar{p}) \operatorname{erfc}(\bar{p})^{1/2}, \quad (54.56)$$

where  $\operatorname{erfc}(\bar{p})$  is the complementary error function defined in eqn. (40.21). The corresponding solution for a paraboloid of revolution, i.e. a needle with a circular cross-section, was obtained by Ivantsov as

$$\bar{\alpha} = \bar{p} \exp(\bar{p}) \int_{\bar{p}}^{\infty} (1/\eta) \exp(-\eta) d\eta. \quad (54.57)$$

Horvay and Cahn give the general solution for the elliptical paraboloid which we can write in the form

$$\bar{\alpha} = f(\bar{p}), \quad (54.58)$$

where  $f(\bar{p})$  has limiting values (54.56) and (54.57) when the aspect ratio of the elliptical cross-section is  $\infty$  and 1 respectively.

In contrast to eqn. (54.52), the growth rate given by (54.56) for small supersaturations is now proportional to the square of the supersaturation, or equivalently to the square of the supercooling  $\Delta T^-$  from the equilibrium condition. Horvay and Cahn express the general result for the elliptical paraboloids in the form

$$Yr \propto \bar{\alpha}^z \propto (\Delta T^-)^z, \quad (54.59)$$

where  $z$  is a slowly varying function of  $\Delta T^-$ . The Zener theory gives  $z = 1$ , whilst the Ivantsov theory gives  $z \sim 2$  for a flat platelet and  $\sim 1.2$  for a needle with a circular cross-section.

Neglect of the Gibbs–Thomson effect would not be significant if the radius of curvature at the tip of a needle or plate were large in comparison with the critical radius of eqn. (54.37) or (54.55). However, measurements show that this is frequently not so, and there must therefore be a variation in  $c^*$  over the surface of the dendrite. This variation will in turn modify the steady-state shape of the particle, an effect first treated by Temkin (1960). Bolling and Tiller (1961) proposed that a reasonable approximation would be to retain the parabolic shape with an assumed constant concentration at the interface of  $c_i^*$  appropriate to the tip of radius  $r$ . This simply means that eqn. (54.58) becomes

$$\bar{\alpha}_{\text{eff}} = f(\bar{p})$$

or

$$\bar{\alpha} = f(\bar{p}) \{1 + (r_c/r)(\bar{\alpha} / f(\bar{p}))\}. \quad (54.60)$$

Trivedi (1970a) has given an exact solution for the diffusion problem which allows the composition to vary along the interface of a plate but assumes that the shape remains that of a parabolic cylinder. His growth equation may be written in the form

$$\bar{\alpha} = f(\bar{p}) \{1 + (r_c/r) \bar{\alpha} S_2(\bar{p})\}, \quad (54.61)$$

where  $S_2(\bar{p})$  is a function of the Péclet number  $\bar{p}$  which Trivedi and Pound (1969) estimated as  $3[2\bar{p}(1+2\bar{p})]^{-1}$ . Trivedi's exact solution leads to a complex expression for  $S_2(\bar{p})$  with numerical values varying from  $\sim 10$  at  $\bar{p} = 0.07$  to  $\sim 0.45$  at  $\bar{p} = 10$ .

Bolling and Tiller also considered that interface processes should not be neglected in dendritic growth, and the theory of morphological stability described above emphasized the possible significant of interface kinetics. If we assume that a finite free energy change is required to move the interface and that the velocity is linearly proportional to this driving force (eqn. 53.9), we can suppose that this results in a deviation  $\Delta c^B$  of the interface composition from the equilibrium value which is additional to that required by the Gibbs–Thomson effect. The composition at the interface is thus

$$c^{\alpha, B} = c_{\infty}^{\alpha} + \Delta c_r^{\alpha} + \Delta c^B, \quad (54.62)$$

and since the free energy charge on crossing the boundary is in first approximation linearly proportional to  $\Delta c^B$ , we can also write this equation (see eqn. (22.38))

$$c^{\alpha, B} = c_{\infty}^{\alpha}(1 + \Gamma/r) + \Upsilon/M^c, \quad (54.63)$$

where  $M^c$  is the boundary mobility per unit concentration difference. If this expression is used in the Zener theory, eqn. (54.43) is modified to

$$\bar{\alpha}_{\text{eff}} = \bar{\alpha}\{1 - (r_c/r) - (\Upsilon/\Upsilon_c)\} = 2C_{14}\bar{\rho}, \quad (54.64)$$

where  $\Upsilon_c$  is defined as the velocity of a flat interface in the limit that the growth is entirely interface-controlled, i.e.

$$\Upsilon_c = M^c(c^m - c_{\infty}^{\alpha}). \quad (54.65)$$

However, it is the normal velocity of the interface which is proportional to  $\Delta c^B$  and so for constant velocity of a parabolic dendrite the interface driving force  $\Delta c^B$  must also vary along the surface. This complication is included in Trivedi's treatment and the final form which replaces eqn. (54.61) is

$$\bar{\alpha} = f(\bar{\rho})\{1 + (r_c/r)\bar{\alpha}S_2(\bar{\rho}) + (\Upsilon/\Upsilon_c)\bar{\alpha}S_1(\bar{\rho})\}, \quad (54.66)$$

where  $S_1(\bar{\rho})$  has a range of values similar to that of  $S_2(\bar{\rho})$ . In effect, the function  $S_2(\bar{\rho}) \cdot f(\bar{\rho})$  corrects for the variation in composition due to changing curvature along the interface and  $S_1(\bar{\rho}) \cdot f(\bar{\rho})$  similarly corrects for the required variation of  $\Delta c^B$  due to the changing orientation of the interface.

Trivedi (1970b) also obtained an exact solution for the paraboloid of revolution. The growth equation has the same form as eqn. (54.66), but  $f(\bar{\rho})$  is given by (54.57) and  $S_1(\bar{\rho})$  and  $S_2(\bar{\rho})$  are replaced by different functions with roughly the same range of values. Presumably this form of the growth equation also applies to the general elliptical paraboloid with  $f(\bar{\rho})$  given by the Horvay and Cahn isoconcentrate solution. We must remember that the solutions are exact only within the framework of the imposed boundary conditions (just as the Ham and Ivantsov solutions are "exact" for simpler boundary conditions) and the actual variation of concentration over the interface will lead to a deviation from a parabolic shape. According to Trivedi, this effect is not serious provided the tip radius is greater than 2 or 3 times  $r_c$ .

For a given shape and value of  $\bar{\alpha}$  there is an infinity of possible growth rates, corresponding to different tip radii. The observation that plates and needles grow at a constant rate indicates a steady-state diffusion field with a constant tip radius which is stable against fluctuations. Zener's assumption that the plate or needle evolves to the form corresponding to maximum velocity has no detailed justification, but until recently it was a working hypothesis, to be tested against experiment.

The maximum velocity at given  $\bar{\alpha}$  is obtained from eqn. (54.66) by putting  $\partial Y/\partial r = 0$ . The value of  $r$  which corresponds to maximum velocity now depends on both  $\bar{\alpha}$  and on a dimensionless parameter

$$\left. \begin{aligned} q &= M^c(c^m - c_\infty^z)r^c/2D && \text{(plates),} \\ q^1 &= 2q && \text{(needles),} \end{aligned} \right\} \quad (54.67)$$

where the interface kinetics term is negligible for large  $q$  and predominates for small  $q$ . The ratio  $r/r_c$  varies considerably with  $\bar{\alpha}$  and  $q$ , and a striking result is that for moderate supersaturations,  $r/r_c$  can exceed 10. The larger radius of curvature leads to smaller maximum velocities, and it is predicted that needles grow faster than plates and should thus be the preferred growth form. In fact, needles form only at small supersaturations and Trivedi attributed this to the variation of the diffusion coefficient with concentration (see p. 494) or to the lower strain energy of a plate.

Tests of the growth theories by simultaneous measurement of tip geometry and velocity were first attempted for solid dendrites by Purdy (1971) but the most influential work has undoubtedly been the experiments of Glicksman *et al.* (1976) on the growth of dendrites of succinonitrile from liquid solution. Their results show that whilst the non-isococoncentrate (or non-isothermal) steady state solutions of the diffusion equation are substantially correct, the actual growth rates predicted by the maximum velocity hypothesis are much too large. In a new theoretical treatment, Langer and Müller-Krumbhaar (1978) have succeeded in replacing the maximum velocity hypothesis, long considered to be unsatisfactory, by a stability criterion which is of the form

$$r = 2\pi/\omega_c$$

where  $\omega_c$  is the critical wave-number for a planar instability and is given by eqn. (54.33). Interface kinetics are not considered and the mathematical development is rather complex, but it is shown that in the absence of the Gibbs-Thomson effect, dendritic growth is inherently unstable. The stability criterion reflects both a tendency for the tip to split if the radius increases and for side-branches to form if the radius decreases. The condition may also be expressed in the form

$$Yr^2 = 8\pi^2 D\Gamma = \text{const.}$$

where  $D$  and  $\Gamma$  may have values appropriate to either chemical or thermal diffusivity.

The various theories of diffusional growth described above are based on the assumption that growth is continuous at all parts of the interface. As discussed at the beginning of this chapter, this is expected if the interface is disordered on an atomic scale, but for semi-coherent interfaces there is the possibility that growth takes place, even under diffusion control, only at favourable sites on the interface. Several electron microscopic studies of

the structures of the planar interfaces of precipitate plates, for example, have indicated that such interfaces contain steps of atomic height and have led to the concept of a growth process controlled by the diffusion flux to the steps (for summary, see Aaronson *et al.*, 1970). The interpretation of these observations is rather controversial and they will be discussed in Part II, Chapter 16, but it is useful here to note how the theory of diffusional growth has been modified to allow for a step mechanism.

For a series of linear steps (ledges) of height  $d$  and step density  $k^S$  per unit length, the velocity of the interface in the direction normal to the broad face of the plate is simply

$$\Upsilon = dk^S u, \quad (54.68)$$

where  $u$  is the mean velocity of a step (Cahn *et al.*, 1964). A first approximation to the velocity of step growth may be to treat the step as the edge of a half-plate, and to use the Zener-Hillert equation (54.52) with  $r \simeq d$ . A better theory was developed by Jones and Trivedi (1971) who assumed that an isolated edge is square-ended and that the growth is steady state, i.e. that the solute distribution in a coordinate system moving with the step is independent of the time. In order to simplify the complex boundary conditions, they considered only the limit of low supersaturation (54.14) so that Laplace's equation (54.15) could be used to find the concentration along the step interface. This concentration variation was then used to solve the complete diffusion equation.

Jones and Trivedi define a Peclet number for the step as

$$\bar{p}^S = ud/2D \quad (54.69)$$

[see (54.42)] and also use a parameter  $q^S$  to express the degree of interface control which is identical with (54.67) except that  $r_c$  is replaced by the step height  $d$ . Their solution takes the form

$$u = M^c(c^m - c^a) [1 + 2q^S \alpha(\bar{p}^S) - 2\bar{p}^S \alpha(\bar{p}^S)]^{-1}, \quad (54.70)$$

where  $\alpha(\bar{p}^S)$  is a function which was evaluated numerically and varies from  $\sim 6$  at  $\bar{p}^S = 0.002$  to  $\sim 0.5$  at  $\bar{p}^S = 0.5$ . This equation may be expressed in the same form as (54.66), i.e.

$$\bar{\alpha} = 2\bar{p}^S \alpha(\bar{p}^S) + (\bar{p}^S/q^S) \{1 - 2\bar{p}^S \alpha(\bar{p}^S)\} \quad (54.71)$$

and in the limit of pure diffusion control ( $q^S \rightarrow \infty$ )

$$u = D\bar{\alpha}/d\alpha(\bar{p}^S), \quad (54.72)$$

where  $\alpha(\bar{p}^S)$  is  $\sim 3$  in the range in which the theory is applicable. In the opposite limit of pure interface control, the velocity is given by

$$u = M^c(c^m - c^a). \quad (54.73)$$

Combination of eqns. (54.68) and (54.72) suggests that the diffusion controlled growth rate of a stepped interface will be constant if  $k^S$  is constant, i.e. if there is some mechanism for the formation of new steps at a constant rate. Alternatively, with a spiral growth mechanism as discussed in Part II, Chapter 13, the total step length is kept constant by a topological property of the interface.

The small proportion of the sites at which atoms can be added to the growing  $\beta$ -plate seem physically to imply that the step mechanism under diffusion control must always give a growth rate smaller than that for a disordered interface, but it is noteworthy that after a sufficiently long time, the above theory gives the opposite prediction since the growth rate given by (54.10) decreases with time. This has been discussed by Atkinson *et al.* (1973) who believe this prediction may represent a real result of the greater flux achieved by the point effect in diffusion, and these authors cite some experimental evidence that the growth velocity of a stepped interface sometimes exceeds the calculated velocity for continuous growth. On the other hand, it is not clear that the changes' in boundary conditions produced when a large number of steps sweep successively across an interface have been properly considered; this may set an upper limit given by (54.10), to the velocity of the interface.

The last topic which we shall discuss in this section is that of diffusion in a stress field, which is of considerable interest in view of effects arising from the interaction between solute atoms (or point defects) and dislocations. Consider a dislocated crystal which initially contains a uniform distribution of solute atoms, so that there is a net flow of atoms of one type towards the dislocation lines. There are two distinct types of boundary condition which may be imposed at the dislocation cores. In the first of these, the equilibrium state of the dislocated crystal is a non-uniform solute distribution, with solute atoms all remaining mobile, but building up "atmospheres" around the dislocation lines (see p. 269). (The dislocation lines must be regarded as frozen into the structure.) The second possibility is that the dislocation lines are effectively sinks for solute atoms which are removed from solution. This boundary condition corresponds to the nucleation and growth of precipitate particles on dislocations, but there is no reason why this process should not be preceded by a stage of segregation to dislocations within the parent matrix.

The migration of solute atoms to dislocations is controlled by a drift flow resulting from the interaction energy  $W_i$  of an individual solute atom with a dislocation and by an opposing diffusion flow. The appropriate flux equation is

$$\mathbf{I} = -D \nabla c - (Dc/kT) \nabla W_i, \quad (54.74)$$

where Einstein's relation (41.15) between the diffusion coefficient and the mobility has been used to give the mean drift velocity of an atom. Equation (54.74) is thus valid only under conditions such that  $\nabla W_i$  is sufficiently small for the drift velocity to have the same activation energy as that needed for diffusion, and it may not apply close to the dislocation core. The simple elastic model gives  $W_i = -(C_{16} \sin \theta)/r$  for an edge dislocation (eqn. (30.24)), where  $C_{16}$  is a constant and  $r, \theta$  are the polar coordinates of the atom relative to the dislocation. The equipotential lines are thus circles passing through the centres of the dislocations, and the lines of force normal to these are also circles defined by  $(\cos \theta)/r = \text{const}$ . Cottrell and Bilby assumed that in the early stages of ageing only the drift flow need be considered, so that on the average the atoms move along these equipotential lines until they reach the position of maximum binding near the centre of the dislocation line.

Consider a particular flow line of radius  $R$ . The mean gradient of  $W_i$  on this line is approximately  $dW_i/dR = -C_{16}/R^2$ , and hence the atoms may be considered to move along this circular line of force with a constant velocity  $(C_{16}D/kT)/R^2$ . An atom on average will move

a distance of order  $R$  in travelling round the flow line, and this will take a time of about  $R^3 kT / C_{16} D$ . Correspondingly, after a time  $t$  has elapsed all flow lines with radii less than  $(C_{16} D t / kT)^{1/3}$  have been completely drained of solute atoms.

It is now possible to write an expression for the number of solute atoms which arrive at the dislocation in a time interval  $dt$  by integrating over all active flow lines, that is for all  $R$  greater than the critical radius which has just ceased to operate at time  $t$ . Qualitatively, it is obvious that the volume of crystal supplying solute atoms to unit length of the dislocation increases as  $R^2$  and hence as  $(C_{16} D t / kT)^{2/3}$ .

In a more accurate treatment, the differential equation

$$\frac{\partial c}{\partial t} = \frac{C_{16} D}{k T r^2} \left[ \sin \theta \frac{\partial c}{\partial r} - \frac{\cos \theta}{r} \frac{\partial c}{\partial \theta} \right] \quad (54.75)$$

obtained from (54.74) by neglecting the  $D \nabla c$  terms has to be solved subject to the boundary condition  $c = 0$  at  $r = 0$ . The solution is discontinuous across the curve

$$k T r^3 \left\{ \frac{1}{2} \pi - \theta - \frac{1}{2} \sin 2\theta \right\} / 2 \cos^3 \theta = C_{16} D t \quad (54.76)$$

with  $c = 0$  inside this curve and  $c = c^m$  outside the curve. The total number of impurity atoms removed from solution to the dislocation is thus equal to  $c^m$  times the area enclosed by (54.76) and this gives the original Cottrell–Bilby result

$$N(t) = 3(\pi/2)^{1/2} c^m (C_{16} D t / k T)^{2/3}. \quad (54.78)$$

Bullough and Newman (1962) pointed out that it is not possible to control the core boundary condition in this treatment, i.e. it is necessary to postulate an ideal sink at  $r = 0$ . They refer to this situation as a “Cottrell atmosphere” and distinguish it from a true Maxwellian atmosphere which is a distribution of solute atoms

$$c = c' \exp(-W_i / kT), \quad (54.79)$$

where  $c(x_1, x_2, x_3)$  depends only on  $W_i(x_1, x_2, x_3)$  and  $c' \simeq c^m$ . A Maxwellian atmosphere is the equilibrium configuration if the solubility limit is nowhere exceeded, but it may only be obtained kinetically when the diffusion flow term is included in the flux equation. Bullough and Newman’s usage of Cottrell atmosphere for the particular situation when a precipitation type of boundary condition is applied at the dislocation core (or rather an undefined perfect sink is postulated at the core) is not universal; as noted on p. 272, Cottrell originally introduced the concepts of a Maxwellian atmosphere and a condensed atmosphere, and any preferred distribution of solute atoms around a dislocation is often referred to as a Cottrell atmosphere.

An alternative to the above treatment of the drift flow is to drop the angular terms in the interaction energy and to write  $W_i = -C_{16}/r$ ; this should be equally valid when applied to measurements which are averaged over many dislocations. The governing equation now becomes

$$\partial c / \partial t = (C_{16} D / k T r^2) \{ (\partial c / \partial r) - (c/r) \} \quad (54.80)$$

which has the solution

$$c(r, t) = c^m r / \{r^3 + (3C_{16}Dt/kT)\}^{1/3}. \quad (54.81)$$

This solution is physically more acceptable than the solution of (54.75) since the concentration field (54.81) contains no discontinuities. The number of solute atoms removed to the dislocation in time  $t$  is now given by

$$N(t) = 3^{2/3}\pi c^m (C_{16}Dt/kT)^{2/3} \quad (54.82)$$

which differs from (54.78) only by a small numerical factor. Thus the apparently large change from circular flow lines to radial flow lines makes very little difference to the overall kinetics of the initial segregation.

A related problem which has assumed some importance in the annealing of slightly impure materials after fast neutron irradiation is the segregation of solute atoms to an assembly of small dislocation loops, rather than to a quasi-linear set of dislocations. Bullough, Stanley and Williams (1968) have shown that the assumption of pure drift flow in this case gives an initial dependence of the fraction of available solute which segregates to the dislocation loops as

$$\zeta(t) = 0.38\pi\varrho_1(6\pi r^2 C_{16}Dt)^{3/5}, \quad (54.83)$$

where  $\varrho_1$  is the density of loops of mean radius  $r$ . Thus this modification of the dislocation geometry changes the time dependence of the drift flow from  $t^{2/3}$  to  $t^{3/5}$ .

As the dislocation begins to draw solute from the surrounding region, concentration gradients will be established in the matrix, and the net flux of solute atoms at any point will then consist of contributions from both the drift flow and the normal diffusion flow. This was clearly recognized in the original paper of Cottrell and Bilby, but it is by no means obvious at what stage in the segregation the diffusion flow becomes significant. Further consideration of this problem is postponed to Section 58, but we may mention here that later work has shown that the proportionality of  $c$  to  $t^{2/3}$  should be valid only for the very early stages of any segregation process, and should not include the range of times over which observable precipitation takes place.

## 55. LINEAR GROWTH OF DUPLEX REGIONS

In the two previous sections we have discussed linear growth in transitions which do not involve a composition change, and parabolic growth in transitions which are diffusion controlled. A transformation involving long-range transport of atoms, however, can take place in such a way that the mean composition of the transformed regions is equal to that of the untransformed matrix. The conditions for linear growth are then satisfied, the composition of the untransformed matrix remaining uniform except near the boundary of a growing product region. This concept covers eutectoidal decompositions and precipitation reactions of the discontinuous type, in both of which the transformed regions are duplex and usually consist of one or more colonies of parallel lamellar crystals of the product phases. The transformed regions generally originate on grain boundaries and grow hemispherically into one of the grains; they are often called nodules. We shall frequently adopt

a terminology due to Turnbull and Treafis (1955) and refer to “cells” when we wish to specify transformed regions of duplex structure originating from a single nucleus. All of the crystals of one type in a cell are approximately parallel and have the same lattice orientation; a growing hemispherical nodule may consist of one or more sets of cells. It is believed that often each cell consists of two interwoven single crystals, one of which continually develops new branches to give the lamellar microstructure, and evidence for this is available in some transformations (Hillert, 1962). However, Tu and Turnbull (1969) found the tin lamellae from a lead–tin alloy not to be interconnected after dissolution of the lead matrix, and they suggest that independent nucleation may make a large contribution to the multiplication of lamellae.

The cells have an incoherent boundary with the untransformed matrix, even when one of the equilibrium phases has the same structure as the parent phase. The two types of reaction which are important in practice are shown schematically in Fig. 11.5. In the first

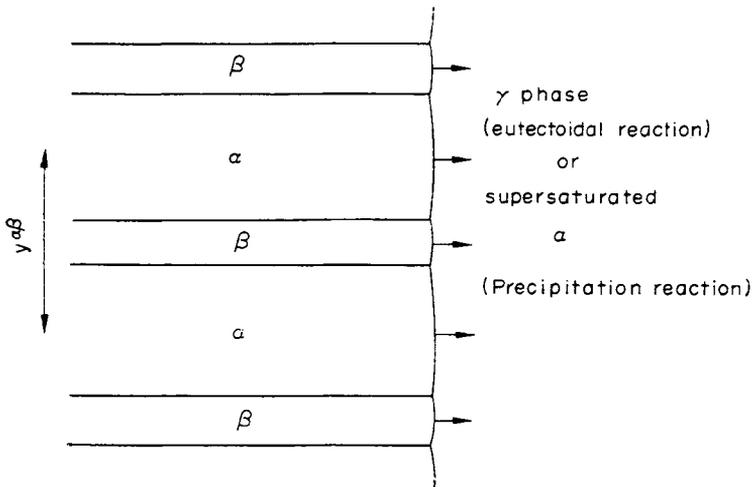


FIG. 11.5. The growth of a duplex lamellar aggregate, or “cell”.

of these, a supersaturated  $\alpha$  solution precipitates a new phase  $\beta$ , and the equilibrium state is  $\alpha + \beta$ . The lamellae of a cell consist of alternate parallel crystals of  $\alpha$  and  $\beta$ . The  $\alpha$  has the same structure as the matrix, but a different orientation, as well as a different composition. This means that the  $\alpha$  matrix and  $\beta$  matrix regions of the cell boundary are both incoherent. In the second reaction, a single phase,  $\gamma$ , decomposes into two new phases,  $\alpha$  and  $\beta$ , and the duplex  $\alpha + \beta$  cell has an incoherent boundary with the  $\gamma$  matrix.

In this section we discuss theories of growth for these duplex product regions. It should be noted that concepts such as nucleation rate and growth rate may be applied to the cell as a whole, as well as to individual crystals within the cell. As the cell grows, the individual plates maintain an approximately constant spacing; this must require the branching of existing plates, or the nucleation of new plates, or both processes. A distinction is sometimes made between the edgewise extension of a set of plates, and the sideways growth which requires “lamellation”—i.e. either nucleation or branching. The growth theories we discuss here

are concerned entirely with edgewise growth. In some cases it has been shown directly that sideways growth takes place at a similar rate to edgewise growth, and other experiments show that either increasing or decreasing the temperature of reaction produces an immediate corresponding change in the spacing of the lamellae, so that lamellation is not a limiting process in the growth (Cahn, 1959).

Mehl (1938) noted that the growth of a lamellar aggregate is a steady-state process, and he suggested that the growth rate should be proportional to both the diffusion coefficient and to an effective concentration gradient in the parent phase. This assumption that solute segregation is obtained by volume diffusion in the matrix has been made in the majority of papers on this kind of growth process, but the possibility of diffusion in the product phases has also been discussed, and the importance of the diffusion short circuit provided by the incoherent boundary of the aggregate has been increasingly recognized. In all theoretical treatments, the solution of the diffusion equation leads to an expression for the product of the interface velocity and either the interlamellar spacing or the square of this spacing, and some additional physical condition is thus required to fix these two quantities separately. Although various principles have been suggested, and will be discussed later in this section, there is no general agreement on this part of the theory, and rigorous arguments in favour of any of the imposed conditions have not been discovered. The growth process is so complex, and there are so many unspecified variables, that it may be more informative to test the experimentally measured growth rates, spacings and diffusion coefficients for self-consistency than to attempt to predict the growth rates and spacings from first principles.

We consider first the situation when the solute segregation is due entirely to diffusion in the matrix. A concentration gradient will be produced because the parent phase is enriched in solute immediately ahead of the solute-poor regions of product ( $\alpha$  in Fig. 11.5) and depleted in solute just ahead of the solute-rich regions ( $\beta$  in Fig. 11.5). Diffusion then transfers atoms normal to the lamellae and parallel to the cell boundary. The process is illustrated in Fig. 11.6(a), and the possible variation of composition along lines through the centres of the plates and normal to the interface is shown schematically in Fig. 11.6(b). This figure also serves to define some concentration terms which we need in the subsequent discussion and which are independent of the assumption of volume diffusion in the parent phase. Within the lamellae, the concentrations of solute at the mid-points of each plate immediately adjacent to the boundary with the matrix are written  $c^\alpha$  and  $c^\beta$ , and the corresponding compositions in the matrix immediately adjacent to these mid-points are written  $c^{m\alpha}$  and  $c^{m\beta}$  respectively. For growth controlled by parent phase diffusion, the compositions within the product lamellae will be uniform; in most treatments it is assumed that the products have their equilibrium compositions at the reaction temperature considered, but we emphasize that there is no *a priori* reason for such an assumption. When the growth is controlled by the diffusion rate, the assumption of local equilibrium at the interface should be valid. For the particular case of a eutectoidal reaction, Hultgren (1938) and others have suggested that the compositions of the  $\gamma$  phase adjacent to the mid-points of the  $\alpha$  and  $\beta$  plates should be those obtained by extrapolation of the  $\alpha + \gamma/\gamma$  and  $\beta + \gamma/\gamma$  boundaries to the reaction temperature. In effect, this assumption gives a maximum value to the composition difference  $c^{m\alpha} - c^{m\beta}$  for a binary eutectoidal system.

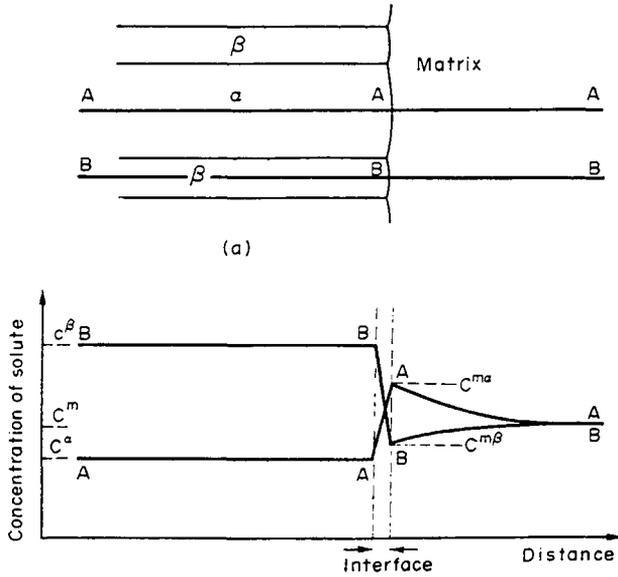


FIG. 11.6. Schematic variation of solute composition for cellular growth controlled by diffusion in the parent phase.

The diffusion gradient and the growth rate may be assumed to be determined by the interlamellar spacing,  $y^{\alpha\beta}$ , since the diffusion effects the segregation of the solute between the  $\alpha$  and  $\beta$  phases; this result is common to all theories in which growth is attributed to volume diffusion. The growth rate for given solute separation does not continue to increase as  $y^{\alpha\beta}$  decreases, however, since the net driving force for the reaction also decreases; this effect is identical with that already considered on p. 501. When the cell boundary advances through a distance  $dx$ , the chemical free energy released in the volume swept out by unit length of each pair of plates cannot exceed  $-(y^{\alpha\beta} \Delta g' / v) dx$ , where  $v$  is the volume per atom and  $\Delta g'$  is the change in free energy per atom produced by transformation of the whole structure to the equilibrium  $\alpha + \beta$  state, neglecting the energy of the interfaces. To avoid the complications of volume changes, we assume that the volume per atom is approximately equal in all three phases. The increase of surface energy in the volume element is  $2\sigma^{\alpha\beta} dx$ , and in the limiting case when all the available driving force is needed to supply this interfacial energy

$$(y^{\alpha\beta})_{\min} = -2\sigma^{\alpha\beta}v/\Delta g'. \tag{55.1}$$

A smaller interlamellar spacing is not possible since there would then be an increase in free energy as the cell grows. If the growth process leads to  $\alpha$  and  $\beta$  phases of non-equilibrium compositions, the chemical driving force may be written  $-P^f \Delta g'$ , where  $P^f$  specifies the fractional amount of chemical free energy released. If the growth velocity increases at fixed spacing, the degree of segregation achieved, and hence  $P^f$ , will decrease.

The minimum spacing for a fixed degree of segregation is  $(y^{\alpha\beta})_{\min}/P^f$ , and corresponds to zero growth velocity, that is to an equilibrium between the product phases and the matrix.

For a planar interface, this would require the chemical potential of any component to be equal in all three phases, a condition which is clearly not satisfied except at the thermodynamic transition temperature of a eutectoidal alloy. However, a metastable equilibrium may be possible because of the Gibbs–Thomson effect when the edges of the lamellae are curved. The three phases must have constant compositions, and there must therefore be constant curvatures at the edges of the plates. This specifies the interface shape, and ensures three-phase equilibrium at  $\alpha$ – $\beta$  interface junctions, but for complete equilibrium the static condition (35.1) has also to be satisfied.

In treating the growth problem the assumption of local equilibrium at the interface and at the three phase junctions is also frequently made. The spacing must now be greater than the minimum value, and the matrix composition must vary along the interface, since there can be no discontinuity in composition at the triple junction, and hence the curvature of the interface must vary if the compositions of the product plates are uniform. The calculation of growth rate, interlamellar spacing, variation of composition and shape of the interface thus becomes a very complex problem even with specific assumptions about rate controlling mechanisms.

We now turn to a consideration of the diffusion problem in the case when diffusion in the parent phase produces the segregation. The earliest theory to give independent expressions for both growth rate and spacing was that of Zener (1946), and we find it convenient to give an outline of his argument here, although it will be evident that it contains many approximations or inconsistencies. Zener made no attempt to calculate the growth rate by a rigorous solution of the diffusion equation, but he assumed that eqn. (54.4) may be applied to each plate and hence to the whole interface. This is clearly unsatisfactory, since the equation was derived for conditions in which the concentration gradients are normal to the interface, whereas in cellular growth the gradients parallel to the interface are more important. The assumption is equivalent to one implicit in an earlier discussion by Mehl, namely that the complex diffusion field may be replaced by a simple effective concentration gradient, and we may tentatively accept it if we recognize that the meanings of the composition terms are rather uncertain. Presumably  $(c^\beta - c^\alpha)$  in eqn. (54.4) has to be replaced by either  $(c^\beta - c^{m\beta})$  or by  $(c^{m\alpha} - c^\alpha)$ , depending on whether we choose to apply the equation to the centre parts of the  $\beta$  or  $\alpha$  plates. The equation requires that  $\dot{\gamma}$  is proportional to a characteristic concentration difference in the matrix,  $\Delta c$ , and Zener supposed that  $\Delta c$  is linearly dependent on the curvature of the  $\alpha$ -matrix or  $\beta$ -matrix boundary respectively. This is effectively equivalent to the assumption that the diffusion flux, and hence  $\Delta c$ , is proportional to the net change in free energy produced by growth. Both chemical and surface terms must be included in this net free energy change, which may thus be regarded as the effective driving force for growth; in the case that the equilibrium segregation is achieved, it is given by

$$-\Delta g^{\gamma} = -\Delta g^t - 2(\sigma^{\alpha\beta}v/y^{\alpha\beta}). \quad (55.2)$$

If we now make  $\Delta c$  proportional to  $-\Delta g^{\gamma}$  we obtain

$$\Delta c = [1 - \{(y^{\alpha\beta})_{\min}/y^{\alpha\beta}\}] \Delta c_{\infty}, \quad (55.3)$$

where  $\Delta c_\infty$  is the concentration difference for a plane boundary. The diffusion distance of eqn. (54.4) is taken to be proportional to  $y^{\alpha\beta}$  and is written  $C_{17}y^{\alpha\beta}$ , where  $C_{17}$  is a constant. The equation for the growth rate of a  $\beta$  plate then becomes

$$\frac{\Upsilon}{D} = \frac{\Delta c_\infty}{(c^\beta - c^{m\beta})} \frac{(y^{\alpha\beta} - (y^{\alpha\beta})_{\min})}{C_{17}(y^{\alpha\beta})^2}, \quad (55.4)$$

and the maximum value of  $\Upsilon$  corresponds to  $y^{\alpha\beta} = 2(y^{\alpha\beta})_{\min}$ . This maximum growth velocity is given by

$$\Upsilon = \frac{D}{2y^D} \frac{\Delta c_\infty}{(c^\beta - c^{m\beta})} = \frac{D}{4C_{17}(y^{\alpha\beta})_{\min}} \frac{\Delta c_\infty}{(c^\beta - c^{m\beta})}. \quad (55.5)$$

Zener's condition for maximum growth rate thus requires that one-half of the driving force is transformed reversibly into the surface energy of the  $\alpha$ - $\beta$  interfaces. This treatment is, of course, virtually identical with that on p. 499.

From eqn. (55.1) we may estimate the variation of  $(y_{\alpha\beta})_{\min}$  with temperature, and hence also the variation of  $y^{\alpha\beta}$ . Using a common approximation, though perhaps with less justification than on p. 444, we write  $\Delta g' = \Delta h' (T^E - T)/T^E$ , where  $T^E$  is the thermodynamic transition temperature. This gives the temperature variation

$$y^{\alpha\beta} \propto (\Delta T^-)^{-1}, \quad (55.6)$$

where  $\Delta T^- = T^E - T$  is the supercooling. In the eutectoidal reaction in steels, the spacing of the plates in pearlite apparently fits this law, but the slope of the experimental line cannot be reconciled with the theory since it leads to an unreasonably large value for  $\sigma^{\alpha\beta}$ . Data for both the kinds of reaction illustrated in Fig. 11.5 indicates that  $y^{\alpha\beta}$  is appreciably larger than  $2(y^{\alpha\beta})_{\min}$ .

If a numerical value for the growth rate is to be deduced, the composition terms and the diffusion distance have to be specified, and here there is considerable confusion in the literature. As already noted, Zener applied eqn. (54.4) to the growth of individual plates of the aggregate, and he considered specifically the case of cementite plates in pearlite. If we call these plates the  $\beta$  plates, the composition term in the denominator becomes  $c^\beta - c^{m\beta}$ , as in eqn. (55.4), and  $y^D$  was rather arbitrarily taken to be equal to the thickness of the  $\beta$  plate, so that  $C_{17}$  is the fractional width of these plates. Zener further assumed that the concentration difference  $\Delta c$  is equal to  $c^{m\alpha} - c^{m\beta}$ , and he used the Hultgren extrapolation to obtain these quantities. He recognized that this treatment would have to give equal growth velocities to both  $\alpha$  and  $\beta$  plates, and gave this as his reason for making  $y^D$  equal to the thickness of only one of these plates when the composition terms in the denominator are expressed in the form of  $c^\beta - c^\alpha$ . In a later review of the theory, Fisher (1950) wrote the growth rate as

$$\Upsilon = \frac{2D}{y^{\alpha\beta}} \frac{(c^{m\alpha} - c^{m\beta})}{c^{m\alpha}}, \quad (55.7)$$

which is nearly the equivalent of Zener's equation applied to the growth of ferrite ( $\alpha$ ) plates if it is assumed that  $c^\alpha$  is neglected in comparison with  $c^{m\alpha}$ , and that  $y^D = \frac{1}{2}y^{\alpha\beta}$ . The difficulties of reconciling the diffusion distances and composition terms illustrate the limitations

of this semiquantitative approach, but if eqn. (54.4) is to be used, it would seem sensible for reasons of symmetry to take the diffusion distance as proportional to  $y^{\alpha\beta}$  and the composition term in the denominator as  $(c^\beta - c^\alpha)$  for both phases. We shall now consider the attempts which have been made to obtain a rigorous solution to the diffusion equation, and we shall find that these compositions do indeed define a convenient growth parameter.

Consider the edgewise growth of a set of parallel lamellae when all the diffusion takes place in the parent phase. Using an origin in the moving boundary, we have to solve eqn. (54.45) in order to obtain a steady-state growth process. The complete solution may be obtained by separating variables as before, and is now periodic in the  $x_2$  direction; as first shown by Brandt (1945), it may be written as an infinite series

$$c - c^m = \sum_{n=0}^{\infty} A_n \exp(-\lambda_n x_1) \cos(2\pi n x_2 / y^{\alpha\beta}), \quad (55.8)$$

where the continuous variable  $b$  of eqn. (54.49) has been replaced by  $b_n = 2\pi n / y^{\alpha\beta}$ . The corresponding values of the coefficients  $\lambda$  are

$$\lambda_n = (\Upsilon/2D) [1 + \{1 + (4\pi n D / \Upsilon y^{\alpha\beta})^2\}^{1/2}]. \quad (55.9)$$

In a multi-component system, a solution of form (55.8) exists for each component,  $c^m$  being the initial concentration of that component in the metastable matrix, and the values of  $A_n$  for that component being determined by the boundary conditions. In the present application, the boundary conditions are completely compatible with the existence of a steady-state solution, but unfortunately our knowledge of these conditions is insufficient to determine the coefficients  $A_n$ . If we suppose that the composition is uniform within the  $\alpha$  and  $\beta$  plates, the steady-state solution imposes a condition that the mean concentration over any plane normal to the growth direction in the product shall be the original concentration of that element  $c^m$ , so that

$$y^\alpha c^\alpha + y^\beta c^\beta = y^{\alpha\beta} c^m, \quad (55.10)$$

where  $y^\alpha, y^\beta$  are the mean widths of the individual  $\alpha$  and  $\beta$  plates respectively. Since  $y^{\alpha\beta} = y^\alpha + y^\beta$ , this condition can be written

$$y^\alpha (c^m - c^\alpha) = y^\beta (c^\beta - c^m) = (y^\alpha y^\beta / y^{\alpha\beta}) (c^\beta - c^\alpha). \quad (55.11)$$

In the more general case where the composition in the product phases is not taken to be constant in the  $x_2$  direction, eqn. (55.10) would have to be replaced by

$$\int_{y^\alpha} c(\alpha) dx_2 + \int_{y^\beta} c(\beta) dx_2 = y^{\alpha\beta} c^m. \quad (55.12)$$

Although the mean composition in the product must equal that in the original matrix, this is not necessarily true of the diffusion zone ahead of the growing interface. The average concentration over any plane of constant  $x_1$  in the matrix phase is given by

$$\bar{c}(x_1) = (1/y^{\alpha\beta}) \int_0^{y^{\alpha\beta}} c dx_2 = c^m + A_0 \exp(-\Upsilon x_1 / D). \quad (55.13)$$

There will thus be a zone ahead of the interface which is enriched or depleted in the component considered, except when  $A_0$  for this component is zero. If there is local equilibrium at the interface,  $A_0$  will be zero for some particular matrix composition at each reaction temperature; in the case of a eutectoidal reaction, this matrix composition will be represented by a line in the equilibrium diagram beginning at the eutectoid point. When a depleted or enriched region exists, its extent in the  $x_1$  direction for a particular component will be increased by a high diffusion coefficient, and decreased by a high interface velocity; it is clear from eqn. (55.13) that we may use  $D/\dot{\gamma}$  as a measure of the thickness of the region.

By retaining only the first two terms of the expansion in eqn. (55.8), Brandt obtained an expression for the growth rate in the form

$$\dot{\gamma} = (2\pi D/\dot{\gamma}^{\alpha\beta})\alpha^m \quad (55.14)$$

in which  $D$  refers to the rate-controlling diffusion process and  $\alpha^m$  is a rather complex homogeneous function of the various concentrations shown in Fig. 11.6. The factor  $2\pi$  is introduced purely for convenience. The parameter  $\alpha^m$  may be compared to the growth parameters used in eqns. (54.8)–(54.13) to characterize growth under parabolic conditions. When  $\alpha^m$  is small in comparison with unity, it is possible to show (Cahn and Hagel, 1962) that

$$\alpha^m = \frac{c^{m\alpha} - c^{m\beta}}{2(c^\beta - c^\alpha)}, \quad (55.15)$$

which are just the composition terms which we should expect to find in a “natural” parameter for this type of growth. Combination of eqns. (55.14) and (55.15) now gives an expression for the growth rate which is very similar to that of Zener, and which is independent of the original composition  $c^m$ . The more accurate expression given by Brandt does include a small dependence on  $c^m$ , but the error in neglecting this is less significant than is that resulting from the retention of only two terms in (55.8).

We can now use eqns. (55.11) and (55.15) to see how the growth rate, spacing, diffusion coefficient, and degree of segregation achieved in the product are interrelated. When  $D$  is small or the product  $\dot{\gamma}y^{\alpha\beta}$  is large,  $\alpha^m$  is relatively large. There is an upper limit to  $(c^{m\alpha} - c^{m\beta})$  which corresponds to the assumption of local equilibrium at the interface, so that  $(c^\beta - c^\alpha)$  must become small in these circumstances. In the other extreme,  $\alpha^m$  is small when  $D$  is large or the product  $\dot{\gamma}y^{\alpha\beta}$  is small; since  $(c^\beta - c^\alpha)$  cannot exceed the difference given by the equilibrium phases, it follows that  $(c^{m\alpha} - c^{m\beta})$  must then become small, so that the composition gradients along the interface are reduced. In most treatments it has simply been assumed that the equilibrium phases are produced by the growth process; the realization that this need not be so is largely due to Cahn (1959) and greatly complicates the theory of the growth. Thermodynamic arguments tell us only that the structures and compositions of the product phases must be such as to produce a net decrease in free energy; within this limitation, the boundary could move slowly, producing a high degree of solute separation, or rapidly, producing less solute segregation. This degree of freedom is applicable to a fixed value of interlamellar spacing, and implies that the growth rate may be dependent on some additional kinetic parameter, such as the mobility of the interface. Some authors dispute this dilemma,

regarding it as self-evident that equilibrium conditions obtain at the interface, including the static condition for the equilibrium of the interphase boundaries at the triple junctions, but there does not seem to be any satisfactory way of proving the necessity of such an equilibrium in a kinetic situation.

Even if the segregation achieved is known or may be derived, the theory of the growth still contains a further degree of freedom, which is needed to fix either the spacing or the growth velocity. We have already described Zener's assumption that the spacing is that which corresponds to maximum growth rate. This yields absolute expressions for growth rate and spacing when combined with the Gibbs-Thomson condition, which reduces the effective composition difference available for the diffusion path in the matrix. However, although the assumption of maximum growth rate is intuitively attractive, it is difficult to give any rigorous justification for it, and Cahn comments that it is not obvious why some other physical quantity such as the rate of production of entropy should not be maximized. In his theory of cellular growth by interface diffusion, Cahn suggested that the spacing should be fixed by minimizing the rate of decrease of free energy; this is discussed below. An alternative suggestion (Kirkaldy, 1958, 1962) is that the rate of entropy production should be minimized; this is based on arguments derived from the theory of the thermodynamics of the steady state, as described in Section 15. Theories based on any of these variational principles are rather unconvincing. The important physical condition is presumably that the spacing, interface shape, compositions, etc. are stable against fluctuations of all types (see below).

We have already noted that the Zener spacing is not observed under experimental conditions, which frequently correspond to  $y^{\alpha\beta} \sim 5(y^{\alpha\beta})_{\min}$ . There is an equally large discrepancy in computed growth rates, although this should not perhaps be taken too seriously because of the approximations made in Brandt's treatment of the diffusion problem. A more accurate solution of the diffusion equation was obtained by Hillert (1957), using the results of eqns. (55.8), (55.9), and (55.10). In order to obtain values for the coefficients  $A_n$ , he assumed that the interface is so plane that  $\exp(-\lambda_n x_1)$  is approximately unity along the whole front. Assuming equilibrium conditions as defined by the Gibbs-Thomson relations and the surface free energies, he deduced a very reasonable approximate shape for the interface in the case of pearlite, and Fourier coefficients given by

$$A_n = \frac{\lambda_n (y^{\alpha\beta})^2}{2\pi^3 n^3 D} (c^\beta - c^\alpha) \sin(n\pi y^\alpha / y^{\alpha\beta}) \quad (n > 0), \quad (55.16)$$

and this yields growth velocity

$$\frac{\dot{\gamma}}{D} = \pi^3 \frac{y^\alpha y^\beta}{(y^{\alpha\beta})^2 \sum_1^\infty (1/n^3) \sin^2(n\pi y^\alpha / y^{\alpha\beta})} \frac{2\alpha^m}{y^{\alpha\beta}}, \quad (55.17)$$

where  $\alpha^m$  is the parameter of eqn. (55.15). This equation differs from Zener's equation, or from (55.14) only by a factor dependent on the ratio of the thicknesses of the  $\alpha$  and  $\beta$  plates, i.e. on the original composition  $c^m$ .

Hillert makes Zener's assumption that the parameter  $\alpha^m$  is equal to the maximum value of this parameter  $\alpha^\infty$ , obtained from the equilibrium diagram, multiplied by  $1 - [(y^{\alpha\beta})_{\min} / y^{\alpha\beta}]$

to take care of the effects of curvature. He thus obtains an expression for maximum growth rate which is of the same form as eqn. (55.5). When applied to experimental results for many eutectoidal and precipitation reactions, all of the equations we have derived give growth rates which are very much slower than those observed, that is, they would require coefficients of diffusion orders of magnitude greater than the experimental coefficients of diffusion in the parent phase. The evidence thus shows that the Zener spacing is often not observed and that volume diffusion in the parent frequently does not effect the segregation.

We shall now consider briefly the possibility that the main diffusion paths are in the product rather than in the parent. This seems an attractive possibility for pearlite formation, as pointed out by Fisher (1950), since the diffusion rate of carbon is about 100 times larger in ferrite than it is in austenite. Fisher wrote the solute flux through the ferrite-matrix boundary as  $\Upsilon c^m$  (apparently neglecting  $c^\alpha$  in comparison with  $c^m$ ) and he gave the growth rate as

$$\Upsilon = (2D/y^{\alpha\beta})(\Delta c/c^m), \quad (55.18)$$

where  $D$  is now the diffusion coefficient in the  $\alpha$ ,  $\Delta c$  is the difference in concentration in the  $\alpha$  phase adjacent to the matrix and to the  $\beta$  phase respectively, and the diffusion distance has been taken as  $\frac{1}{2}y^{\alpha\beta}$ . A more rigorous treatment gives similar results. The diffusion equation (54.45) yields the solution

$$c - c^\alpha = \sum_{n=-\infty}^0 A_n \exp(-\lambda_n x_1) (\cos 2\pi n x_2 / y^{\alpha\beta}) \quad (55.19)$$

with

$$\lambda_n = (\Upsilon/2D) [1 - \{1 + (4\pi n D / \Upsilon y^{\alpha\beta})^2\}^{1/2}] \quad (n < 0),$$

which may be interpreted in a similar way to that already discussed for the case of parent phase diffusion. Diffusion in the product must involve composition gradients in both phases, which is not too probable, or else the thickness of the  $\beta$  plate must decrease towards zero as the interface is approached. Although such tapering or ending of the  $\beta$  plates is occasionally observed (Darken and Fisher, 1962), it does not seem to be generally found in sufficiently rapidly quenched specimens. Moreover, although the diffusion coefficient in one of the product phases may be appreciably larger than in the matrix, as in the example just quoted, the available composition difference for diffusion is also smaller. Thus it does not appear likely that any appreciable fraction of the solute segregation is achieved by volume diffusion in the region of product immediately behind the interface. High growth rates can then only be explained by some form of diffusion short circuit, and it now seems certain that in many transformations the dominant diffusion process is along the cell boundary and not through the parent or product lattices.

The use of the cell boundary as a diffusion short circuit was suggested by several workers, and an attempt at a quantitative treatment was first made by Turnbull (1955). The incoherent boundary is particularly effective in a reaction of this nature, since it sweeps through the matrix as the cell grows, and is favourably orientated for achieving segregation in the required direction. Since the activation energy for cell boundary diffusion will be appreciably

less than that for lattice diffusion, the relative importance of the processes may change with temperature, boundary diffusion becoming dominant as the temperature is lowered.

In Turnbull's theory of cell growth, eqn. (55.18) is modified to give

$$\Upsilon = \{(D^B \delta^B / (y^{\alpha\beta})^2) \{(c^m - c^\alpha) / c^m\}, \quad (55.20)$$

where  $D^B$  is the diffusion coefficient in the boundary and  $\delta^B$  is the thickness of the boundary. Equation (55.20) contains Fisher's approximate treatment of the concentration terms and a re-evaluation by Aaronson and Liu (1968) leads to the replacements of the last brackets by  $4(c^\beta - c^\alpha) / (c^\beta - c^m) \simeq 4$  when (54.14) applies. Cahn and Hagel (1962) use a slightly different formulation, writing instead of eqn. (55.14)

$$\alpha^B = \Upsilon (y^{\alpha\beta})^2 / 4\pi^2 D^B \delta^B \quad (55.21)$$

since the thickness of the band of sideways diffusion has changed by approximately  $2\pi\delta^B / y^{\alpha\beta}$ . This follows since the diffusion in the parent phase takes place over a band of thickness  $1/\lambda_1$ . When the value of  $\alpha^m$  given by eqn. (55.14) is much smaller than unity,  $1/\lambda_1 = y^{\alpha\beta} / 2\pi$  and the thickness of the diffusion zone is much smaller than the extent of the depleted or enriched zone  $D/\Upsilon$ . If  $\alpha^B$  is much greater than unity, the diffusion band is equal to  $D/\Upsilon$  and smaller than  $y^{\alpha\beta} / 2\pi$ ; little segregation then occurs by volume diffusion. If segregation is achieved by a combination of volume and interface diffusion, the growth rate spacing and composition terms will be related by equating the sum of  $\alpha^m$  and  $\alpha^B$  to the value given by eqn. (55.15). The dominant diffusion mechanism depends on the ratio of the diffusion coefficients, the condition for boundary diffusion being the more important, being

$$D^B / D > y^{\alpha\beta} / 2\pi\delta^B. \quad (55.22)$$

A theory of growth when the boundary mechanism is operative is due to Cahn (1959). He pointed out that with boundary diffusion it is impossible to achieve the equilibrium segregation at any non-zero value of the growth rate, since there must be a composition gradient in the boundary itself, and hence in the  $\alpha$  and  $\beta$  plates, unless the growth rate tends to zero. This means that  $P^f$  (p. 511) must always be less than unity, and the net free energy decrease of the reaction becomes

$$\Delta g^I = P^f \Delta g^f + 2\sigma^{\alpha\beta} v / y^{\alpha\beta} \quad (55.23)$$

in place of (55.2). For fixed  $y^{\alpha\beta}$ , there is a minimum value of  $P^f$ , and a corresponding maximum growth rate,  $\Upsilon_{\max}$ , which gives  $\Delta g^I = 0$ . Equally for fixed  $P^f$ , there is a minimum spacing of  $1/P^f$  times the minimum spacing in eqn. (55.1).

Suppose we fix the spacing and consider the variation in the amount of segregation with boundary velocity. We then encounter the difficulty discussed on p. 515 that any value of  $\Upsilon$  smaller than  $\Upsilon_{\max}$  is thermodynamically possible. A full specification of the growth process now requires a second kinetic parameter in addition to the diffusion coefficient, and the choice of this is not obvious. Cahn assumed that the boundary may be treated in the same way as an ordinary boundary (Section 53) and by analogy with eqn. (53.9), the growth rate

is regarded as the product of a mobility and a free energy driving force. Using the net driving force  $\Delta g^x$ , the growth rate becomes

$$\Upsilon = -M^B \Delta g^x, \quad (55.24)$$

where  $M^B$  is the boundary mobility and  $\Delta g^x$  is given by eqn. (55.23). Combination of this equation with the diffusion equations (55.21) and (55.15) then enables the growth rate and degree of segregation to be determined if the spacing is known.

Some physical principle is still needed to fix the spacing itself. In Zener's theory the spacing was chosen to maximize the growth rate. Cahn adopted the hypothesis that the spacing maximizes the decrease in free energy  $\Delta g^x$ , and points out that the spacing which achieves this also maximizes the growth rate  $\Upsilon$  and the rate of decrease of free energy per unit area of cell boundary. This is thus a very reasonable assumption although justification in a more fundamental manner seems desirable. In Zener's treatment the maxima in the above three quantities occur at quite different spacings, and it is not obvious which should be chosen.

The permissible variation in segregation achieved by cellular growth has quite different forms for precipitation and eutectoidal reactions. This is evident from Fig. 11.7. which

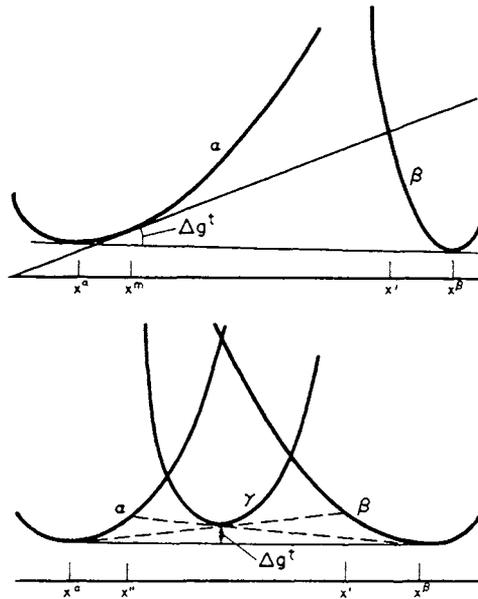


FIG. 11.7. Schematic free energy vs. composition curves for eutectoidal and precipitation reactions (after Cahn, 1959).

shows hypothetical free energy curves for the two kinds of reaction. For the precipitation reaction, a lowering of free energy results from the formation of  $\beta$  phase of any composition richer in solute than  $x'$ , even if the amount of  $\beta$  phase formed is so small that the  $\alpha$  phase has a composition almost identical with that of the original matrix composition  $x^m$ . Thus a small amount of segregation is quite feasible. For the eutectoidal reaction, the composition of the  $\alpha$  phase must be less than  $x''$ , and that of the  $\beta$  phase greater than  $x'$  if the free energy

is to be reduced. This means that neither phase can approximate to the matrix composition, and a large fraction of the equilibrium segregation must be accomplished as the cells grow.

To put the above considerations into quantitative form it is necessary to solve the diffusion equation. Cahn has done this, using an idealized model for each type of reaction. The diffusion is assumed to be confined to a plane boundary of thickness  $\delta^B$  and diffusion coefficient  $D^B$ , independent of concentration. The concentration of solute in the boundary is  $c^B(x_2)$ , where  $x_2$  is a coordinate along the boundary normal to the lamellae. The corresponding concentration in the product regions of  $\alpha$  and  $\beta$  plate is written  $c^P(x_2)$ . Equating the diffusion flow along the boundary to the flow due to the motion of the boundary, in order to obtain a steady state, we then find

$$D^B \delta^B (d^2 c^B / dx_2^2) + \Upsilon (c^m - c^P) = 0, \quad (55.25)$$

which is the equivalent in this model of the volume diffusion equation (54.45).

In solving this equation, Cahn assumes that the  $\alpha$  and  $\beta$  phases both have their equilibrium concentrations at the  $\alpha/\beta$  boundaries. For a precipitation reaction, the concentration of the solute in the  $\alpha$  phase is taken to be proportional to the concentration in the boundary, so that  $c^P(x_2)/c^B(x_2) = C_{18}$ . The origin is taken at the mid-point of an  $\alpha$  plate, and if the supersaturation is small, the  $\beta$  plates are so thin that  $c^P(x_2) = c^\alpha$  at  $x_2 = \frac{1}{2} y^{\alpha\beta}$ . The solution to eqn. (55.25) with this boundary condition can then be written

$$\frac{c^m - c^P}{c^m - c^\alpha} = \frac{\cosh\{(\alpha')^{1/2} x_2 / y^{\alpha\beta}\}}{\cosh\{\frac{1}{2}(\alpha')^{1/2}\}}, \quad (55.26)$$

where the parameter  $\alpha'$  is given by

$$\alpha' = C_{18} \Upsilon (y^{\alpha\beta})^2 / D^B \delta^B = (C_{18} / 4\pi^2) \alpha^B. \quad (55.27)$$

In terms of this parameter, the fraction of excess solute which is precipitated as the cell grows is given by

$$Q^t = \{(2/(\alpha')^{1/2}) \tanh\{\frac{1}{2}(\alpha')^{1/2}\}\}. \quad (55.28)$$

If the free energy curve has a parabolic variation with composition, as is true for dilute solutions, the fraction of the chemical free energy released by the growth is given by

$$P^t = \{3/(\alpha')^{1/2}\} \tanh\{\frac{1}{2}(\alpha')^{1/2}\} - \frac{1}{2} \operatorname{sech}^2\{\frac{1}{2}(\alpha')^{1/2}\}. \quad (55.29)$$

These equations both contain the growth rate and the spacing. An expression for  $\alpha'$  in terms of independent quantities is obtained by maximizing  $-\Delta g^t$  in eqn. (55.23), and using eqn. (55.24). This gives

$$\Delta g^t / \Delta g^t = P^t + 2\alpha' (dP^t / d\alpha')$$

assuming  $P^t$  to be a function of  $\alpha'$  only, and leads to a relation

$$\frac{(\alpha')^3 (dP^t / d\alpha')^2}{\alpha' + 2\alpha' (dP^t / d\alpha')} = - \frac{C_{18} M^B (\sigma^{\alpha\beta})^2 v^2}{D^B \delta^B \Delta g^t} = \beta^t, \quad (55.30)$$

which implicitly gives  $\alpha^t$  in terms of a parameter  $\beta^t$  which may be evaluated independently. Figure 11.8 shows the way in which  $\log \alpha^t$ ,  $P^t$  and  $Q^t$  vary with  $\beta^t$ , increasing  $\beta^t$  representing conditions in which the boundary mobility becomes larger in relation to the diffusion coefficient.

Figure 11.8. also includes plots of the quantities  $R^t$  and  $(P^t - R^t)$ .  $R^t$  is the fraction of the chemical driving force which is converted to surface energy, and is given by

$$\begin{aligned} R^t &= -(2\sigma^{\alpha\beta}v/y^{\alpha\beta} \Delta g^t) = (y^{\alpha\beta})_{\min}/y^{\alpha\beta} \\ &= -2\alpha^t(dP^t/d\alpha^t). \end{aligned} \tag{55.31}$$

In Zener's theory,  $R^t = \frac{1}{2}$ , but as may be seen from the figure, Cahn's model allows it to

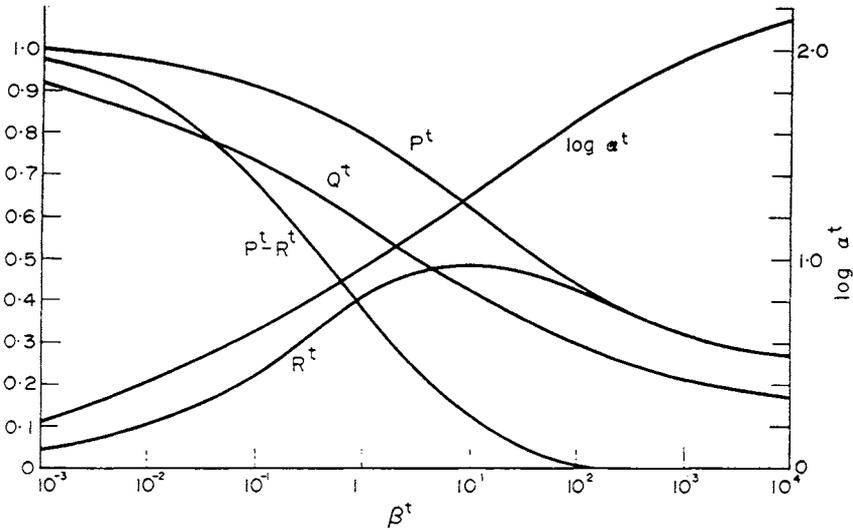


FIG. 11.8. Variation of growth parameters with the parameter  $\beta^t$  of (55.30) for a discontinuous precipitation reaction in which growth depends on boundary diffusion (after Cahn, 1959).

have any value from 0 to a maximum of 0.49. The quantity  $(P^t - R^t)$  is the driving force available to exert pressure on the boundary.

Examination of Fig. 11.8 shows that for a highly mobile boundary with a low diffusion coefficient,  $P^t = R^t$ , and all the free energy released is converted to surface energy. The growth is almost reversible, and the boundary velocity is governed only by the thermodynamic requirement that  $P^t - R^t$  shall be positive. The boundary velocity is thus  $\Upsilon = \Upsilon_{\max}$ . Conversely, if the diffusion coefficient is relatively high, and the mobility low, the proportion of the chemical energy released is much higher. Since the boundary can now achieve segregation efficiently, the spacing can also be large, and  $R^t$  becomes very small. A large spacing permits more of the chemical driving force to be utilized in moving the rather immobile boundary.

The other idealized case considered by Cahn is that of a symmetrical eutectoid, with the eutectoid composition at  $x_B = \frac{1}{2}$  and the two new phases having compositions  $x^\alpha$  and  $x^\beta =$

$1 - x^\alpha$ . We shall not discuss this in detail; the results are shown in Fig. 11.9 in terms of a parameter  $\beta^t = -\beta^t \Delta g^t / C_{19}$ , where  $C_{19}$  is a constant specifying the (assumed parabolic) free energy curves. The parameter  $\beta^t$  has some advantage over  $\beta^t$  in being nearly independent of temperature. Figure 11.9 is plotted for a particular assumption,  $-\Delta g^t / C_{19} = 10^{-2}$ , so that it may readily be compared with Fig. 11.8 for the precipitation reaction. It is seen that for high diffusion rate and low mobility, the sets of curves are very similar, but this is not true of high mobilities and low diffusion rates. As mentioned earlier,  $Q^t$  is able to

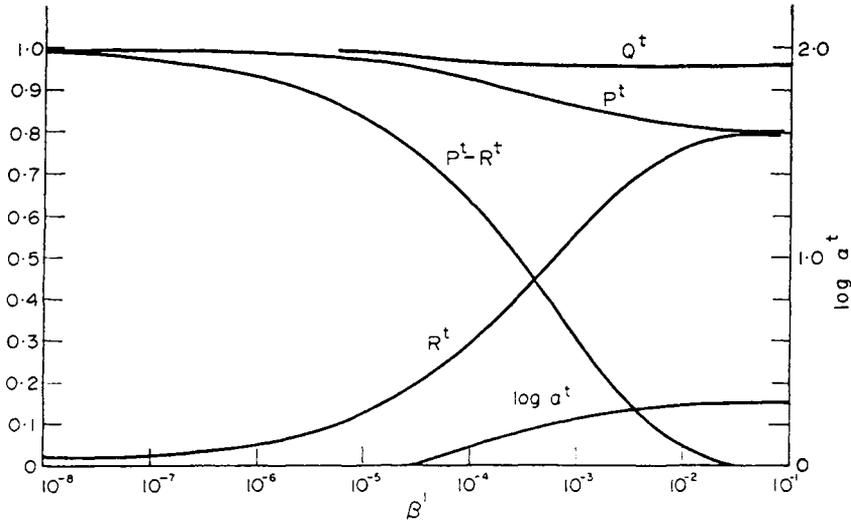


FIG. 11.9. Variation of growth parameters with the parameter  $\beta^t$  for a eutectoidal reaction in which growth depends on boundary diffusion (after Cahn, 1959).

vary little in this type of reaction, and a fine spacing is needed in order to accomplish the required segregation when the mobility is high. The requirement that  $(P^t - R^t)$  be positive can then only be met if about 80% of the available free energy is converted into surface energy. Thus  $P^t$  and  $R^t$  both tend to  $\sim 0.8$ , instead of to the much lower value in Fig. 11.8. This high value of  $R^t$  means that the spacing in a eutectoidal reaction can be much finer even than Zener's prediction, approaching a limit of  $y^{\alpha\beta} \approx 1.25(y^{\alpha\beta})_{\min}$ .

Cahn's theory contains some unverified assumptions but is clearly more satisfactory than earlier treatments. Although the detailed results in Figs. 11.8 and 11.9 depend on the properties of the idealized models, the general trends will be common to all reactions of the two types provided the assumptions are correct. Modern techniques allow the theory to be tested, since  $\Upsilon$ ,  $y^{\alpha\beta}$  and  $Q^t$  may all be measured. Liu and Aaronson (1968) conclude that both the modified Turnbull theory and the Cahn theory give reasonable values for  $D^B$  when used in conjunction with experimental results on lead-tin alloys, but that predicted values of  $\sigma^{\alpha\beta}$  are much too high and temperature sensitive. Speich (1968) found his experimental results on iron-zinc alloys implied that the interface velocity depends on the cube of the driving force, so that eqn. (55.24) is replaced by

$$\Upsilon = -M^1(\Delta g^t)^3. \quad (55.32)$$

The modified Cahn theory which resulted from this growth has predicted spacings  $\gamma^{\alpha\beta}$  within a factor of two of those observed.

Some authors (e.g. Hillert, 1968) have criticized the Turnbull and Cahn theories of discontinuous precipitation because they do not consider explicitly the forces acting on the  $\alpha'-\alpha$  boundary and thus give no clear indication why the boundary should move. The difficulty is not encountered in a rather special model (Shapiro and Kirkaldy, 1968) which is based on the assumption that there is a metastable monoeutectoid in systems exhibiting discontinuous precipitation, but this seems unsatisfactory as a general condition. A more general way of avoiding the difficulty (Hillert, 1968; Sundquist, 1973) is to use a model in which the grain boundary is treated as a separate phase.

The most recent theories of lamellar growth (Sundquist 1968, 1973) are based on the hypothesis that the solute segregation is produced by boundary diffusion, but local equilibrium at the interface is assumed for eutectoidal decomposition, whereas a deviation from equilibrium is required in discontinuous precipitation. The interface spacing in these theories is determined by stability criteria rather than by optimization principles.

In his theory of eutectoidal growth, which is applied specifically to pearlite in steels, Sundquist (1968) uses the Gibbs-Thomson effect to relate the interface shape with the composition variation along the interface, and he does not introduce a mobility parameter. For a given spacing, solution of the diffusion equation enables the shape and velocity of the interface to be approximately computed. This theory is thus similar to the growth theory developed much earlier by Hillert (1957), but is modified to take account of interface diffusion rather than volume diffusion.

Since the interface is no longer considered to be planar and diffusion is allowed only along the interface, it is convenient to use the interface to define a curvilinear coordinate  $l$  in place of  $x_2$  in the diffusion equation (55.25). The angle between the local direction of  $l$  and  $x_2$  is  $\theta$ , and it follows also that the velocity in the equation is  $\Upsilon_\theta = \Upsilon \cos \theta$  where  $\Upsilon$  is the velocity in the overall growth direction (see p. 503). In this theory the interface diffusion coefficient  $D^B$  is allowed to have separate values  $D^{B\alpha}$  and  $D^{B\beta}$  in the  $\gamma-\alpha$  and  $\gamma-\beta$  interfaces. The interface concentration is assumed to be proportional to the concentration in the austenite adjacent to the interface, and this in turn is governed by the Gibbs-Thomson effect. This gives (cf. eqn. (22.38))

$$c^B = K^B c_\gamma^P = K^B c_\infty^P (1 + \Gamma^P/r), \quad (55.33)$$

where  $P$  means either  $\alpha$  or  $\beta$  depending on whether the  $\alpha-\gamma$  or  $\beta-\gamma$  interface is considered,  $c_\infty^P$  represents the composition of  $\gamma$  in equilibrium with either  $\alpha$  or  $\beta$  at a planar interface,  $r$  is the local radius of curvature, and  $\Gamma^P$  is given by eqn. (22.34) or one of its approximate forms. The diffusion equation (55.25) becomes

$$\Upsilon \cos \theta (c^m - c^P) = -D^B K^B \delta^B \Gamma^P \{\partial^2(1/r)/\partial l^2\}. \quad (55.34)$$

Since  $1/r = d\theta/dl$ , this equation may be put in the form

$$B^P \cos \theta = \partial^3 \theta / \partial l^3. \quad (55.35)$$

Equations (55.33)–(55.35) may each be applied separately to the  $\gamma$ - $\alpha$  and  $\gamma$ - $\beta$  boundaries. However, at the three phase junction,  $c^{\gamma\alpha}$  must equal  $c^{\gamma\beta}$ , so that if the origin of  $l$  is chosen at this point

$$c^{\gamma\alpha}\{1 + \Gamma^\alpha(\partial\theta^\alpha/\partial l)_{l=0}\} = c^{\gamma\beta}\{1 + \Gamma^\beta(\partial\theta^\beta/\partial l)_{l=0}\}. \quad (55.36)$$

There are also two other boundary conditions, namely that  $(\theta^P)_{l=0}$  has values  $\theta_0^P$  which satisfy the condition for local equilibrium of surface forces (eqn. 35.1), and that

$$(\partial^n\theta/\partial l^n)_{l=L^P} = 0 \quad (n = 2, 4, \dots), \quad (55.37)$$

where the superscript  $P$ , as before, identifies  $\alpha$  or  $\beta$ , and  $L^P$  is the value of  $l$  at the mid-points of either  $\alpha$  or  $\beta$  plates respectively. Clearly, the approximation  $\cos\theta \simeq 1$  requires

$$L^\alpha \simeq \frac{1}{2}y^\alpha, \quad L^\beta \simeq \frac{1}{2}y^\beta,$$

but the exact relation is

$$y^P = 2 \int_0^{L^P} \cos\theta^P dl. \quad (55.38)$$

Condition (55.37) ensures that the interface profiles are symmetrical about the mid-points of the plates.

Sundquist obtained a numerical solution for the interface velocity and shape given by eqn. (55.35) and an approximate analytical solution by setting  $\cos\theta = 1$ . In terms of the following parameters

$$R_D = D^{B\alpha}/D^{B\beta}, \quad R^P = c_\infty^{\gamma P} \Gamma^P \theta_0^P, \quad (55.39)$$

$$U^B = \gamma/D^{B\alpha} K^B \delta^B,$$

the analytical solution is

$$U^B = \frac{3\{c_\infty^{\gamma\beta} - c_\infty^{\gamma\alpha} + (R^\alpha/L^\alpha) - (R^\beta/L^\beta)\}}{\{(c^m - c^\beta)(L^\beta)^2/R_D\} - (c^m - c^\alpha)(L^\alpha)^2}. \quad (55.40)$$

The quantity  $K^B U^B$  thus replaces  $4\pi^2\alpha^B/(\gamma^{\alpha\beta})^2$  of eqn. (55.21).

A plot of  $U^B$  against  $\gamma^{\alpha\beta}$  is negative for small  $\gamma^{\alpha\beta}$ , rises steeply to a positive maximum, and then decreases slowly towards zero when  $y^{\alpha\beta}$  becomes large. Sundquist's numerical calculations showed that this curve is insensitive to  $R_D$  and to the values of  $\theta_0$ . With one set of assumptions ( $R_D = 3.0$ ,  $\theta_0^\alpha = \theta_0^\beta = 30^\circ$ ), he plotted a family of curves for different temperatures and found that the experimental values of  $y^{\alpha\beta}$  fitted the result

$$U^B(y^{\alpha\beta})^3 = \text{const.} \quad (55.41)$$

Since  $y^{\alpha\beta} \propto (\Delta T^-)^{-1}$  it follows that the temperature dependence of the growth velocity is given by

$$Y \propto (\Delta T^-)^3 D^B K^B. \quad (55.42)$$

Assuming that  $D^B$  is given by an equation of type (44.13) and that  $K^B = K_0^B \exp(\Delta h_K/kT)$ , where  $\Delta h_K$  is the heat evolved on depositing a solute atom on the interface, experimental results on the variation of velocity with temperature may now be used to derive an apparent

activation energy which represents the difference between the activation energy for diffusion and  $\Delta h_K$ . The results for pearlite give a value for this energy of the order of the activation energy for lattice diffusion of carbon in austenite (or twice that for carbon in ferrite), and this is clearly not consistent with the assumption of growth controlled by interface diffusion. Sundquist attributes the discrepancy to the effects of impurities in reducing the growth rate.

As in the case of the theories of growth controlled by volume diffusion, the observed interlamellar spacings are of the order of five times the spacing which would give maximum growth rate. Following Cahn (1959) and Jackson and Hunt (1966), Sundquist shows that spacings smaller than that of maximum growth rate will be unstable and will rapidly converge to the spacing of maximum growth, unwanted lamellae being simply left behind or "grown over". He refers to this as the lower catastrophic limit. Spacings which are very large are also unstable because there is then no steady-state shape to the interface. As shown schematically in Fig. 11.10, the interface doubles back on itself at large spacings, and at a

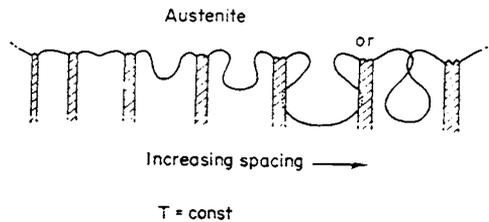


FIG. 11.10. Schematic representation of pearlite-austenite, interface shapes as a function of interlamellar spacing (all other conditions held constant) (Sundquist, 1968).

certain spacing the maximum negative angle increases rapidly with further increase of spacing. This leads to the formation of deep recesses in the centre of the  $\alpha$  plates, either exposing new  $\beta$ - $\gamma$  interfaces which leads to branching, or enclosing volumes of  $\gamma$  phase which then nucleate new lamellae. In either case, the result is to produce a rapid decrease in spacing down to a limit (the upper catastrophic limit) at which a steady-state solution (velocity constant along the interface) is possible.

Between these two catastrophic limits, spacing changes are slower and continuous and are produced by the motion of lamellar faults (internally terminating lamellae). Thus the theory makes no definite prediction about observed spacings in this region, but the experimental results seem to correspond to spacings near the upper catastrophic limit. The reason for this is suggested to be that pearlite grows as colonies which have overall interfaces which are not planar but are convex towards the austenite. This requires a constant supply of new lamellae to maintain a given spacing and the lamellar fault mechanism may be inadequate for this; hence, growth takes place near the upper limit of stability.

The mechanism for maintaining the spacing is also a prominent feature of the theory of discontinuous precipitation developed by Sundquist (1973). He adapts a theory due to Tu and Turnbull (1967) based on experimental results which indicate that the  $\beta$  particles in an  $\alpha + \beta$  cell growing from  $\alpha'$  are not formed by branching from a singular nucleus but by successive nucleation of favourably orientated regions of  $\alpha$ - $\alpha'$  boundary. In Sundquist's description, this nucleation occurs wherever the interlamellar spacing becomes sufficiently

large (because of the diverging growth of a cell with a curved interface) to produce recesses in the  $\alpha'$ - $\alpha$  boundary. The theory is thus essentially similar to the earlier explanation for the maintenance of the pearlite spacing near the upper limit of stability. The suggested process is shown schematically in Fig. 11.11.

The quantitative theory developed by Sundquist depends on rather detailed assumptions and will not be described here. Four "forces" are considered to act on the  $\alpha'$ - $\alpha$  boundary, namely a curvature (Gibbs-Thomson) force, a solute drag from adsorbed solute which moves with the boundary, the force resulting from departure from local equilibrium (described as a "negative solute drag"), and the intrinsic drag or grain boundary kinetics effect. This last mobility force is included in Cahn's theory but is here considered negligible. The

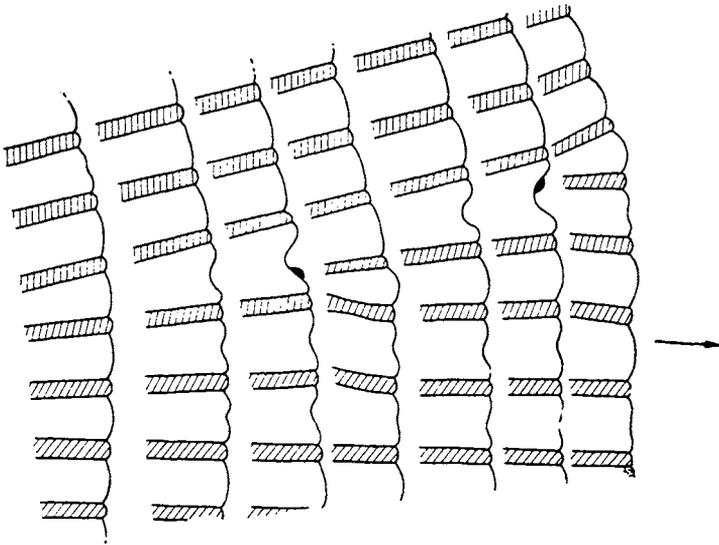


FIG. 11.11. Schematic representation of the proposed mechanism controlling the interlamellar spacing during cellular precipitation (Sundquist, 1973).

solute drag (see Part II, Chapter 19) is treated by means of a model due to Hillert (1968) in which the grain boundary is regarded as a separate phase rather than as a transition region. The results of the theory may be expressed in the form

$$D^B \delta^B K_0^B = \Upsilon (y^a)^2 \bar{Q}^2 / 4(1 - \bar{Q}^4)^2 \cos \theta, \quad (55.43)$$

and this equation should be compared with the corresponding Turnbull and Cahn expressions eqns. ((55.20) and (55.21)) and with Sundquist's theory of eutectoidal growth (eqn. 55.40). In eqn. (55.43)  $K_0$  is a coefficient which depends on the interaction between the solute and the boundary,  $\bar{Q}$  is related to the fractional amount of excess solute precipitated  $Q'$  but is modified to take account of the deviation from the equilibrium value  $c^a$  of the composition of the  $\alpha$  phase adjacent to the  $\beta$  lamellae immediately behind the interface, and  $\theta$  is the average value of the interface inclination.

Sundquist has compared his theory with the experimental results of Liu and Aaronson (1968) and Speich (1958) referred to above and concludes that it is consistent with reasonable values for grain boundary diffusivities and other data. The available experimental results on discontinuous precipitation and eutectoidal reactions are considered more fully in Sections 74 and 77. However, we should note here that as for eutectoidal reactions, the results strongly suggest that the interface velocity is much more than linearly dependent on driving force. A detailed relation between spacing  $\gamma^{\alpha\beta}$  and supercooling  $\Delta T^-$  is not derived, but the experimental results are claimed to be consistent with the criterion shown in Fig. 11.11.

## REFERENCES

- AARONSON, H. I. and LIU, Y. C. (1968) *Scripta metall.* **2**, 1.  
 AARONSON, H. I., LAIRD, C., and KINSMAN, K. R. (1970) *Phase Transformations*, p. 313. A.S.M. Metals Park, Ohio.  
 ATKINSON, C. (1967) *Acta metall.* **15**, 1207; (1968) *Ibid.* **16**, 1019.  
 ATKINSON, C., KINSMAN, K. R. and AARONSON, H. I. (1973) *Scripta metall.* **7**, 1105.  
 BECKER, J. H. (1958) *J. Appl. Phys.* **29**, 110.  
 BOLLING, G. F. and TILLER, W. A. (1961) *J. Appl. Phys.* **32**, 2587.  
 BRANDT, W. H. (1945) *J. Appl. Phys.* **16**, 139.  
 BULLOUGH, R. and NEWMAN, R. C. (1962) *Acta metall.* **10**, 971.  
 BULLOUGH, R. STANLEY, J. T. and WILLIAMS, J. M. (1968) *Metal Sci. J.* **2**, 93.  
 CAHN, J. W. (1959) *Acta metall.* **7**, 20; (1960) *Ibid.* **8**, 556.  
 CAHN, J. W. and HAGEL, W. C. (1962) *Decomposition of Austenite by Diffusional Processes*, p. 131, Interscience, New York.  
 CAHN, J. W., HILLIG, W. B. and SEARS G. W. (1964) *Acta metall.* **12**, 1421.  
 COTTRELL, A. H. and BILBY, B. A. (1949) *Proc. Phys. Soc.* **62**, 49.  
 DARKEN, L. S. and FISHER, R. M. (1962) *Decomposition of Austenite by Diffusional Processes*, p. 249, Interscience, New York.  
 DOREMUS, R. H. (1957) *Acta metall.* **5**, 393.  
 FISHER, J. C. (1950) *Thermodynamics in Physical Metallurgy*, p. 201, American Society for Metals, Cleveland, Ohio.  
 FRANK, F. C. (1950) *Proc. R. Soc. A*, **201**, 586.  
 GLICKSMAN, M. E., SHAEFER, R. J. and AYERS, J. D. (1976) *Metall. Trans. (A)* **7**, 1747.  
 HAM, F. S. (1958) *J. Phys. Chem. Solids* **6**, 335; (1959a) *Q. J. Appl. Math.* **17**, 137; (1959b) *J. Appl. Phys.* **30**, 915; (1959c) *Ibid.* **30**, 1518.  
 HILLERT, M. (1957) *Jernkontorets Ann.* **141**, 757; (1962) *Decomposition of Austenite by Diffusional Processes*, p. 197. Interscience, New York; (1968) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 231, Institute of Metals, London.  
 HILLIG, W. B. and TURNBULL, D. (1956) *J. Chem. Phys.* **24**, 914.  
 HORVAY, G. and CAHN, J. W. (1961) *Acta metall.* **9**, 695.  
 HULTGREN, A. (1938) *Hardenability of Alloy Steels*, p. 55, American Society for Metals, Cleveland, Ohio.  
 IVANTSOV, G. P. (1947) *Dokl. Akad. Nauk. SSSR* **58**, 567.  
 JACKSON, K. A. and HUNT, J. D. (1966) *Trans. Am. Inst. Min. (Metall.) Engrs.* **236**, 1129.  
 JONES, G. J. and TRIVEDI, R. (1971) *J. Appl. Phys.* **42**, 4299.  
 KIRKALDY, J. S. (1958) *Can. J. Phys.* **36**, 907; (1959) *Ibid.* **37**, 739; (1962) *Decomposition of Austenite by Diffusional Processes*, p. 39, Wiley, New York.  
 LANGER, J. S. and MÜLLER-KRUMBHAAR, H. (1978) *Acta metall.* **26**, 1681, 1689, 1697.  
 LIU, Y. C. and AARONSON, H. I. (1968) *Acta metall.* **16**, 1343.  
 MACHLIN, E. S. (1953) *Trans. Am. Inst. Min. (Metall.) Engrs.* **197**, 437.  
 MEHL, R. F. (1938) *Hardenability of Alloy Steels*, p. 1, American Society for Metals, Cleveland, Ohio.  
 MOTT, N. F. (1948) *Proc. Phys. Soc.* **60**, 391.  
 MULLINS, W. W. and SEKERKA, R. F. (1963) *J. Appl. Phys.* **34**, 323; (1964) *Ibid.*, **35**, 444.  
 PAPAPETROU, A. (1935) *Z. Kristallogr.* **92**, 108.

- PHILIP, J. R. (1960) *Aust. J. Phys.* **13**, 1.  
PURDY, G. R. (1971) *Met. Sci. J.*, **5**, 81.  
SHAPIRO, J. M. and KIRKALDY, J. S. (1968) *Acta Met.* **16** 1239.  
SHEWMON, P. G. (1965) *Trans. Met. Soc. AIME* **233**, 736.  
SPEICH, G. R. (1968) *Trans. Met. Soc. AIME* **242**, 1359.  
SUNDQUIST, B. E. (1968) *Acta metall.* **16**, 1413; (1973) *Metall. Trans.* **4**, 1919.  
TEMKIN, D. E. (1960) *Soviet Physics-Doklady* **5**, 609.  
TRIVEDI, R. (1970a) *Metall. Trans.* **1**, 921; (1970b) *Acta metall.* **18**, 287.  
TRIVEDI, R. and POUND, G. M., (1967) *J. Appl. Phys.* **38**, 3569; (1969) *Ibid.* **40**, 4293.  
TU, K. N. and TURNBULL, D. (1967) *Acta metall.* **15**, 369, 1317; (1969) *Ibid.* **17**, 1263.  
TURNBULL, D. (1955) *Acta metall.* **3**, 55.  
TURNBULL, D. and TREAFTIS, H. C. (1955) *Acta metall.* **3**, 43.  
WERT, C. (1949) *J. Appl. Phys.* **20**, 943.  
ZENER, C. (1946) *Trans. Amer. Inst. Min. (Metall.) Engrs.* **167**, 550; (1949) *J. Appl. Phys.* **20**, 950.

## CHAPTER 12

# *Formal Theory of Transformation Kinetics*

### 56. TRANSFORMATIONS NUCLEATED ON GRAIN BOUNDARIES

A brief introduction to the theory of isothermal transformation curves was given in Section 4, and we now attempt to remove some of the restrictive assumptions. We begin by considering transformations which do not involve a change in mean composition, including polymorphic phase changes, single phase processes such as recrystallization, and reactions of the type considered in Section 55. In all these changes, concentration gradients either do not exist or are present only in the immediate vicinity of a boundary, and their extent does not depend on the position of the boundary. Steady-state conditions should then be quickly established, so that growth rates are constant. The conclusion that any dimension of a growing region is a linear function of the time has been verified experimentally for a number of reactions. The theory of Chapter 10 also indicates that under some circumstances a steady-state, time-independent nucleation rate per unit volume of untransformed material should be attained, but there are many reactions for which this will not be true.

An isotropic growth rate  $\Upsilon$  was assumed in Section 4 and we shall continue to use this assumption in most of the subsequent development of this chapter. This is in agreement with experimental evidence that in many transformations the reaction product grows approximately as spherical nodules. The extension to the general case of anisotropic growth is readily made if the shape of the growing region stays constant. We can then represent the growth rate in any direction in terms of the principal growth velocities  $\Upsilon_1, \Upsilon_2, \Upsilon_3$  in three mutually perpendicular directions and the volume of a region originating at time  $\tau$  is

$$\begin{aligned} v_\tau &= \eta \Upsilon_1 \Upsilon_2 \Upsilon_3 (t - \tau)^3 & (t > \tau), \\ v_\tau &= 0 & (t < \tau), \end{aligned}$$

where  $\eta$  is a shape factor. For homogeneous nucleation throughout the body of the assembly, the treatment of Section 4 still applies, and the volume fraction transformed at time  $t$  is

$$\zeta = 1 - \exp \left[ \eta \Upsilon_1 \Upsilon_2 \Upsilon_3 \int_0^t v I (t - \tau)^3 d\tau \right] \quad (56.1)$$

and in particular, for constant  $vI$

$$\zeta = 1 - \exp(-\eta \Upsilon_1 \Upsilon_2 \Upsilon_3 v I t^4 / 4). \quad (56.2)$$

A more general assumption about nucleation is that there are  ${}^vN_0$  nuclei pre-existing at time  $t = 0$ , and in addition there is a subsequent nucleation rate  ${}^vI = C_{20}t^n$ . Substituting into eqn. (56.1), the form of the kinetic law becomes

$$\zeta = 1 - \exp - [\gamma \Upsilon_1 \Upsilon_2 \Upsilon_3 \{ {}^vN_0 t^3 + C_{20} t^{4+n} \}] \quad (56.3)$$

(see also (4.10)).

In view of the conclusions of Section 51, we must next examine the possibility that nucleation occurs only at grain boundary surfaces, grain edges, or grain corners. Johnson and Mehl (1939) gave a treatment of grain boundary nucleation in which they assumed that nodules of reaction product grow only in the grains in which they nucleate, and cannot cross grain boundaries. This assumption was not made in deriving eqn. (56.2), and if it is valid, the previous results for random volume nucleation will have to be modified for transformations in fine grained material. However, the available experimental evidence seems more in favour of the alternative assumption that grain boundaries offer no resistance to growing nodules. If a boundary does stop a transformed region, nucleation on the other side of the boundary at this place is much more probable than at random points along the boundary, and the assumption of continued growth is a fair approximation. We shall thus not give Johnson and Mehl's analysis here, and we shall neglect grain boundary hindrance to growth.

The calculation of the isothermal reaction curves for nuclei forming preferentially at either grain boundaries (where two grains meet) or at edges or corners (three and four grains meeting respectively) is due to Cahn (1956a). In Section 4 we introduced the extended volume, which is the volume of all transformed regions, assuming that each one never stops growing, and that nuclei continue to form in transformed as well as untransformed regions. Now consider any plane surface, of total area  $O$ . The extended area  $O_e^\beta$  is defined as the sum of the areas of intersection of the extended nodules with this plane. Let the area of real transformed regions intersected by this plane be  $O^\beta$ , and consider a small period of time, during which these two quantities change by  $dO_e^\beta$  and  $dO^\beta$ . Then if the intersections which together comprise  $O_e^\beta$  are randomly distributed on the plane, a fraction  $(1 - O^\beta/O)$  of the elements which make up  $dO_e^\beta$  will also contribute to  $dO^\beta$ , so that

$$\left. \begin{aligned} dO^\beta &= (1 - O^\beta/O) dO_e^\beta \\ O_e^\beta/O &= -\ln(1 - O^\beta/O). \end{aligned} \right\} \quad (56.4)$$

This equation, and its derivation, are identical with that given previously for  $V_e^\beta$  (eqn. (4.5)). Clearly, we may also define an extended line intercept  $L_e^\beta$ , which is the sum of the lengths of the intercepts cut off on length  $L$  by the extended volumes. It is related to the sum of the lengths cut off by real transformed regions by

$$L_e^\beta/L = -\ln(1 - L^\beta/L).$$

For simplicity, we assume an isotropic growth rate  $\Upsilon$  in deriving an expression for grain boundary nucleated transformation. A nodule nucleated on a plane boundary at time  $\tau$  will then intersect an arbitrary plane parallel to the boundary and distant  $y$  away from it in a circle. At time  $t$ , the radius of this circle is  $[\Upsilon^2(t - \tau)^2 - y^2]^{1/2}$  when  $\Upsilon(t - \tau) > y$ , and zero

when  $\Upsilon(t-\tau) < y$ . Suppose that  ${}^B I$  is the specific grain boundary nucleation rate per unit area of the boundary (defined operationally in the same way as  ${}^V I$  in Section 4), and let the boundary have area  $O^b$ . Then, at time  $t$ , the extended transformed area intersected by our reference plane is  $O_e^\beta$ , and the contribution to this extended area from regions nucleated between times  $t = \tau, \tau + d\tau$  is

$$\begin{aligned} dO_e^\beta &= \pi O^b {}^B I [\Upsilon^2(t-\tau)^2 - y^2] & [\Upsilon(t-\tau) > y], \\ dO_e^\beta &= 0 & [\Upsilon(t-\tau) < y], \end{aligned}$$

and the whole extended area is

$$O_e^\beta = \int_{\tau=0}^t dO_e^\beta = \pi O^b \int_0^{(t-y/\Upsilon)} \{\Upsilon^2(t-\tau)^2 - y^2\} {}^B I d\tau.$$

This expression can be integrated if  ${}^B I$  is assumed to be constant. Introducing the new variable  $\xi = y/\Upsilon t$  for convenience, we find

$$\begin{aligned} O_e^\beta &= \pi O^b {}^B I \Upsilon^2 t^3 (1 - 3\xi^2 - 2\xi^3)/3 & (\xi < 1), \\ O_e^\beta &= 0 & (\xi > 1). \end{aligned}$$

Since the intersections making up  $O_e^\beta$  are randomly distributed, the true area of transformed material intersected by this plane is related to  $O_e^\beta$  by eqn. (56.4).

We next calculate the total volume of all transformed material originating from this grain boundary, assuming there is no interference with growth by regions originating in other boundaries. By treating  $y$  as a variable, and allowing it to take all values from  $-\infty$  to  $+\infty$ , we find that this volume is given by

$$\begin{aligned} 2 \int_0^\infty O^\beta dy &= 2 \Upsilon t \int_0^1 \{1 - \exp(-O_e^\beta)\} d\xi \\ &= 2 O^b (\Upsilon/{}^B I)^{1/3} f^B(a^B), \end{aligned} \tag{56.5}$$

where

$$a^B = ({}^B I \Upsilon^2)^{1/3} t$$

and

$$f^B(a^B) = a^B \int_0^1 [1 - \exp\{(-\pi/3)(a^B)^3(1 - 3\xi^2 - 2\xi^3)\}] d\xi.$$

Now consider the whole assembly to contain a large number of planar grain boundaries of total area  $O^B = \Sigma O^b$ . Replacing  $O^b$  in (56.5) by  $O^B$  gives the total volume of transformed material nucleated at the boundaries, on the assumption that regions from different boundaries do not interfere. This volume is thus an extended volume in which allowance has been made for mutual impingement of nodules starting from one planar boundary, but not for impingement of regions starting from different boundaries. This extended volume can be related to the true transformed volume if it is assumed that the planar boundaries are themselves randomly distributed in space. Equation (4.5) then applies, and the transformed volume fraction is

$$\zeta = 1 - \exp\{-(b^B)^{-1/3} f^B(a^B)\}, \tag{56.6}$$

where the further abbreviation

$$b^B = {}^B I / \{8({}^v O^B)^3 \Upsilon\} \quad (56.7)$$

has been introduced. The quantity  ${}^v O^B$  is the grain boundary area per unit volume.

The case of edge nucleation is treated similarly. The concept of extended line intercept is used to calculate the total volume of nodules originating from one straight edge, assuming impingement only with other nodules nucleated on the same edge. A random distribution of edges is then considered, in order to allow for impingement of regions nucleated on different edges. Since the calculation is virtually identical with that above, we shall only quote Cahn's final result, which is

$$\zeta = 1 - \exp\{-(b^E)^{-1} f^E(a^E)\}, \quad (56.8)$$

where

$$a^E = ({}^E I \Upsilon)^{1/2} t, \quad b^E = {}^E I / \{2\pi {}^v L^E \Upsilon\}, \quad (56.9)$$

and

$$f^E(a^E) = (a^E)^2 \int_0^1 \xi [1 - \exp\{-(a^E)^2 [(1 - \xi^2)^{1/2} - \xi^2 \ln\{(1 + (1 - \xi^2)^{1/2})/\xi\}]\}] d\xi,$$

and  ${}^E I$  is the specific edge nucleation rate per unit length of edge, and  ${}^v L^E$  the boundary edge length per unit volume.

Finally, there is the possibility that nucleation occurs only at grain corners. This is equivalent to Avrami's assumption of a limited number of randomly distributed nucleation sites, and the transformation eqn. (4.9) is immediately applicable. To facilitate comparison with the results in this section, we write it in the form

$$\zeta = 1 - \exp\{-(b^C)^{-3} f^C(a^C)\}, \quad (56.10)$$

where

$$a^C = {}^C I t, \quad b^C = (3/4\pi {}^v N^C)^{1/3} ({}^C I / \Upsilon) \quad (56.11)$$

and

$$f^C(a^C) = (a^C)^3 - 3(a^C)^2 + 6a^C - 6\{1 - \exp(-a^C)\}.$$

${}^C I$  is the specific corner nucleation rate (i.e. the nucleation rate per corner site) and corresponds to the quantity  $v_1$  used in Section 4.  ${}^v N^C$  is the density of corner sites, and corresponds to the quantity  ${}^v N_0$  of Section 4.

## 57. ANALYSIS OF ISOTHERMAL TRANSFORMATION CURVES

Experimental determinations of some physical quantity such as electrical resistivity, specimen length, or relative intensities of X-ray diffraction lines, enable the  $\zeta-t$  relations to be found with fair accuracy in many transformations. If the theory of Sections 4 and 53 is applicable, the most useful way of analysing the data is usually to plot curves of  $\log \log [1/(1-\zeta)]$  against  $\log t$ . When the general equation (4.11) applies, such curves should be straight lines of slope  $n$ . Figure 12.1 is an example, taken from unpublished work on the transformation from  $\beta$  to  $\alpha$  manganese.

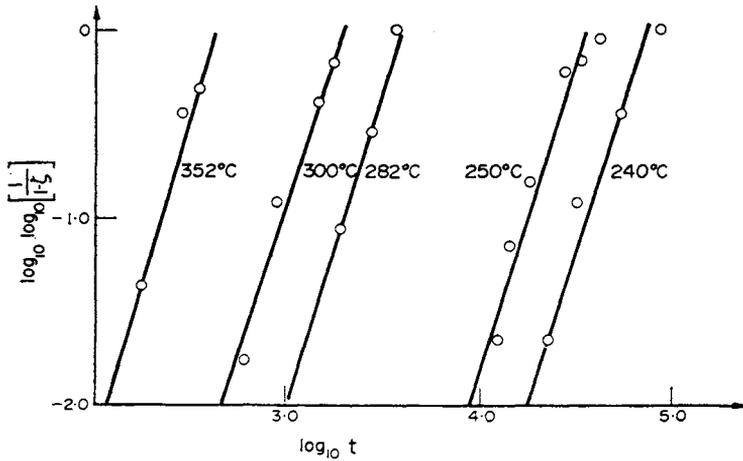


FIG. 12.1. Kinetics of the transformation from  $\beta$  to  $\alpha$  manganese (Husband *et al.*, 1959).

The existence of a straight line relation might be thought to imply random volume nucleation, since the functional dependence in eqns. (56.6), (56.8), and (56.10) cannot be expressed in the simple form (4.11). In fact we shall find this is not so. Consider first that nucleation is on grain boundaries. When  $a^B$  is very small, eqn. (56.6) approaches the limiting form

$$\zeta = 1 - \exp(-\pi vI^B \Upsilon^3 t^4/3), \tag{57.1}$$

where  $vI^B = vO^B B I$  is the grain boundary nucleation rate per unit volume of the assembly. This expression is identical with the equation (4.7) or (56.2)) for random volume nucleation, so that  $\zeta$  depends only on the nucleation rate per unit volume, irrespective of where the nuclei are formed. When  $a^B$  is very large, (56.6) has another limiting form

$$\zeta = 1 - \exp(-2 vO^B \Upsilon t). \tag{57.2}$$

The  $\log \log [1/(1-\zeta)] - \log t$  plot thus consists of two straight lines, of slopes four and one, with an intermediate region over which the slope decreases. From eqn. (56.6) we see that a  $\{\log f^B(a^B)\} - \{\log a^B\}$  plot is equivalent to a  $\{\log \ln[1/(1-\zeta)] + \frac{1}{3} \log b^B\} - \{\log t + \frac{2}{3} \log(^B I \Upsilon^2)\}$  plot. This curve, part of which is shown in Fig. 12.2, is thus a master curve for all grain boundary nucleated reactions, and it is related to an actual  $\{\log \log [1/(1-\zeta)]\} - \{\log t\}$  plot only by two additive constants. The experimental curve for any reaction to which the theory applies should thus fit this curve merely by moving the origin.

The physical explanation for the change in slope in Fig. 12.2. was termed by Cahn "site saturation". It occurs because the nucleation sites on the boundaries are not randomly distributed in the volume, but are concentrated near other nucleation sites. This means that at any stage the fraction of the boundary area transformed is greater than the volume fraction transformed, and because of this, the overall nucleation rate, which depends on the untransformed boundary area, decreases more rapidly than does the untransformed volume. Before saturation occurs, the fact that nuclei are confined to grain boundaries scarcely affects the overall volume fraction transformed, and eqn. (57.1) applies. After satu-

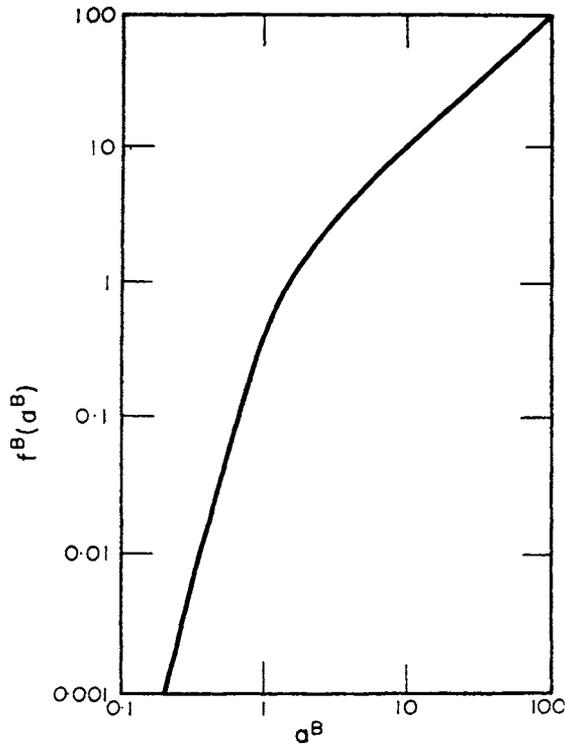


FIG. 12.2. Master curve for transformations which nucleate on grain boundary surfaces (after Cahn, 1956a).

ration, the later stages of the reaction correspond to effectively zero nucleation rate. Equation (57.2) can, in fact, be obtained simply by considering the growth laws of slabs of transformation product spreading into the grains from the grain boundaries.

The above discussion might lead us to expect that the experimental  $\{\log \log [1/(1-\zeta)]\} - \{\log t\}$  curves will show the bend in the middle of Fig. 12.2. However, variation of  $\zeta$  from 0.01 to 0.99, which represents the maximum observable experimental range, only covers a range of 2.7 in the ordinates of Fig. 12.2. It is thus probable that the whole observable range of reaction will correspond to one or other of the straight line regions even when nucleation is confined to the grain boundaries. Site saturation occurs when  $a^B \approx 1$ , i.e. at times  $t \approx 1/({}^B I \Upsilon^2)^{1/3}$ . This is independent of  $vO^B$ , and hence of the grain size, but the grain size is important in determining whether or not this time corresponds to an observable stage of the reaction.

Site saturation will be observed if it occurs when  $\zeta \approx 0.5$ , i.e. when

$$b^B = \{f^B(a^B)/\ln 2\}^3 \quad (57.3)$$

and  $f^B(a^B)$  is given a value  $\sim 1.25$  corresponding to the bend in Fig. 12.2. From (56.7) we see that site saturation occurs at half reaction if

$${}^B I \approx \{125(vO^B)^3 \Upsilon\} / \{8(\ln 2)^3\}.$$

Substituting the approximate value for  $vO^B$  given on p. 334, and writing  $vI^B = vO^B bI$ , this equation becomes

$$vI^B \approx 6 \times 10^3 \Upsilon / (L^B)^4. \tag{57.4}$$

For values of  $vI^B$  smaller than some value near that given by this equation, saturation of nucleation sites will not occur until a late stage of the reaction, and the kinetics are equivalent to those of random volume nucleation. This was realized by Avrami, who pointed out that this treatment also applied to nucleation which is only "locally random". For larger values of  $vI^B$ , saturation occurs early in the reaction. Only for a small critical range, where the condition (57.4) holds almost exactly, should the change of slope be discernible on a  $\{\log \log [1/(1-\zeta)]\}-\{\log t\}$  curve. In the previous chapter we have noted that the nucleation rate changes very rapidly with the degree of supercooling (or superheating) from the thermodynamical transition temperature. It follows that the transition between the two linear functions representing the extremes of Fig. 12.2 will occur in a very small temperature interval. Effectively we may say that there is a critical temperature, defined by the condition (57.4). At temperatures nearer to the thermodynamical transition temperature, the kinetics of isothermal transformation will be indistinguishable from those resulting from random volume nucleation. At temperatures more remote than the critical temperature, the kinetics will suggest that the grain boundaries all constitute nuclei pre-existing at the beginning of the transformation.

Reactions which are nucleated at grain edges or grain corners may be treated in the same way. The master curves, which are effectively plots of  $\{\log \ln [1/(1-\zeta)] + \log b^E\}$  against  $\{\log t + \frac{1}{2} \log {}^E T\}$  and of  $\{\log \ln [1/(1-\zeta)] + 3 \log b^C\}$  against  $\{\log t + \log {}^C T\}$ , are shown in Figs. 12.3 and 12.4. Site saturation occurs when the edges or corners all lie in transformed

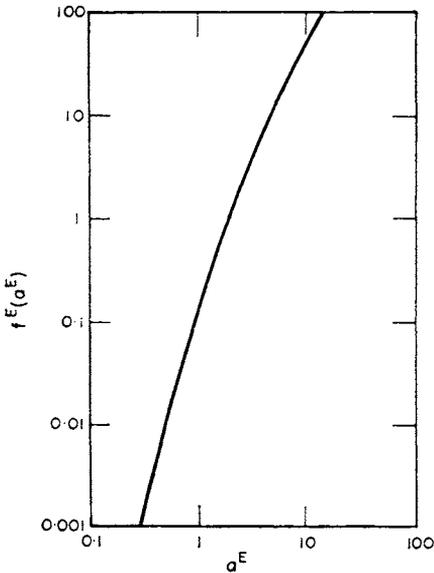


FIG. 12.3. Master curve for transformations which nucleate on grain edges (after Cahn, 1956a).

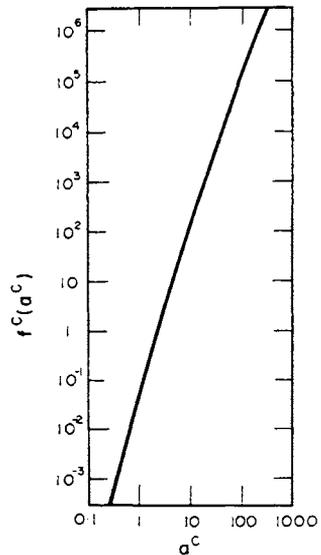


FIG. 12.4. Master curve for transformations which nucleate on grain corners (after Cahn, 1956a).

material. When  $a^E$  or  $a^C$  is small, the equation for  $\zeta$  in both cases again approaches (57.1), so that before saturation is reached, the kinetics are identical with those for random volume nucleation. When  $a^E$  is large, eqn. (56.8) becomes

$$\zeta = 1 - \exp(-\pi vL^E \Upsilon^2 t^2) \quad (57.5)$$

and when  $a^C$  is large, the corresponding limit of (56.10) is

$$\zeta = 1 - \exp(-4\pi vN^C \Upsilon^3 t^3/3) \quad (57.6)$$

(see also eqn. (4.10)), Once again, there is a bend in the transformation curves in the region of  $a^E$  or  $a^C \approx 1$ , but this is less pronounced for edges and still less for corners. Even if the constants are such that site saturation occurs at about half transformation, it is unlikely that the bend in the curve will be detected experimentally; instead a straight line of intermediate slope may be found.

Equation (57.3), and the equivalent expressions for edge and corner nucleation, may be used to determine the dependence of the time to half transformation  $t_{1/2}$ , on the grain size  $L^B$ . Using the appropriate expressions for  $vO^B$  and substituting for  $b^B$

$$({}^B I/\Upsilon)^{1/3} L^B \approx 9.7 f^B(a_{1/2}^B), \quad (57.7)$$

where  $a_{1/2}^B = ({}^B I/\Upsilon)^{1/3} \Upsilon t_{1/2}$ . A plot of  $\log\{9.7 f^B(a^B)\}$  against  $\log a^B$  is thus equivalent to plotting  $\{\log L^B + \frac{1}{3} \log ({}^B I/\Upsilon)\}$  against  $\{\log \Upsilon t_{1/2} + \frac{1}{3} \log ({}^B I/\Upsilon)\}$ . This curve is shown in Fig. 12.5; it differs from Fig. 12.2 only in a change of origin. Provided the quantities  $L^B$  and  $\Upsilon t_{1/2}$  are measured in the same units, any line of unit slope in this figure represents a

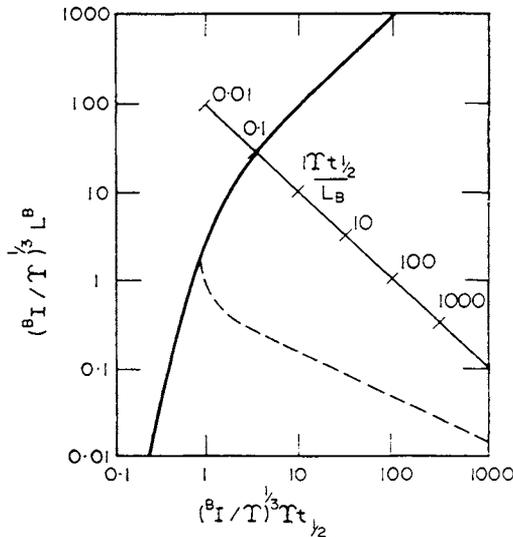


FIG. 12.5. Theoretical curve for half-times of transformations which nucleate on grain boundary surfaces (after Cahn, 1956a). The broken line gives the results of Johnson and Mehl (1939).

set of constant values of  $(Yt_{1/2}/L^B)$  with varying  $(^BI/Y)$ . A scale of  $(Yt_{1/2}/L^B)$  can thus also be drawn. From this, we see that when site saturation has occurred,  $\{(^BI/Y)^{1/3} L^B > \sim 12\}$ :

$$t_{1/2} \simeq 0.1(L^B/Y) \tag{57.8}$$

and that if saturation does not occur, the half-time of the reaction must be greater than  $0.1(L^B/Y)$ . If  $^BI$  and  $Y$  are fixed (at constant temperature),  $t_{1/2}$  decreases as  $L^B$  decreases, but once the grain size is sufficiently small to prevent saturation of nucleation sites by half-transformation time, the decrease in  $t_{1/2}$  is less rapid than the decrease in  $L^B$ . At sufficiently low values of  $(^BI/Y)^{1/2} L^B$ , the curve of Fig. 12.5 has a slope of four as we have seen previously. In this region

$$t_{1/2} \simeq (6/9.7\pi)(L^B/^BIY^3)^{1/4}.$$

This result, expressed in the form  $t_{1/2} \propto (^BIY^3)^{-1/4}$ , is the same as that for random volume nucleation, but the nucleation rate per unit volume,  $^vI$ , is of course independent of grain size for volume nucleation.

The continual increase of  $t_{1/2}$  with  $L^B$  in Fig. 12.5 is simply a result of the decreasing ratio of grain boundary area to volume, since we have assumed that the transformed volume starting from a single nucleus is not restricted by the grain boundaries. The alternative assumption of Johnson and Mehl that transformed regions cannot cross grain boundaries gives an almost identical curve in the limit when the grain size is large. For small grain sizes, however, the amount of transformation per nucleus is limited not by impingement but by the grain boundaries, and is thus proportional to  $(L^B)^3$ . Since the volume nucleation rate only increases as  $(L^B)^{-1}$ , the overall transformation rate decreases as  $(L^B)^2$ . Figure 12.5 also shows the curve obtained from Johnson and Mehl's calculation; it will be seen that there is an optimum grain size of  $L^B \simeq (Y/^BI)^{1/3}$  for rapid transformation, and this corresponds to a transformation time of  $t_{1/2} \simeq (1/^BIY^2)^{1/3}$

Figures equivalent to 12.4 can be obtained for edge and corner nucleation in the same way. For edge nucleation, a curve of  $\log\{8.8[f^E(a^E)]^{1/2}\}$  against  $\log a^E$  is equivalent to plotting  $\{\log L^B + \frac{1}{2} \log(^EI/Y)\}$  against  $\{\log Yt_{1/2} + \frac{1}{2} \log(^EI/Y)\}$ . For corner nucleation, a curve of  $\log\{4.2[f^C(a^C)]^{1/3}\}$  against  $\log a^C$  is equivalent to plotting  $\{\log L^B + \log(^CI/Y)\}$  against  $\{\log Yt_{1/2} + \log(^CI/Y)\}$ . In all cases, site saturation leads to half-times given by (57.8), and very small grain sizes lead to  $t_{1/2} \propto (^vIY^3)^{-1/4}$ . Since  $^vI \propto (L^B)^{-2}$  for edges and  $(L^B)^{-3}$  for corners, we can now write down the functional dependence of  $t_{1/2}$  on  $L^B$  when other parameters are fixed. The results are gathered together in Table VIII.

We may also generalize this result by considering a nucleation rate which is a power function of the time, of the form

$$^vI = C_{20}t^n. \tag{57.9}$$

It is readily seen that  $t_{1/2} \propto (L^B)^m$ , where

$$m = (3-i)/(4+n) \tag{57.10}$$

and  $i$  is the dimensionality of the site. The values of  $i$  for homogeneous nucleation and for nucleation on grain boundaries, edges and corners are 3, 2, 1, and 0 respectively.

TABLE VIII. DEPENDENCE OF TIME TO HALF TRANSFORMATION ON GRAIN SIZE, ASSUMING CONSTANT NUCLEATION RATE  $t_{1/2} \propto (L^B)^m$

Type of nucleation	$m$
Random volume	0
Grain boundary	$\frac{1}{4}$
Grain edge	$\frac{1}{2}$
Grain corner	$\frac{3}{4}$
Site saturation (all types)	1

It follows from this discussion that experimental investigation of the variation of  $t_{1/2}$  with grain size may give useful information about the type of nucleation which is operative. Few investigations of this type appear to have been made.

In a real transformation, it is possible that several kinds of site are active at the same time, all with different characteristic nucleation rates. When all sites are unsaturated, the rate law is approximately that for random nucleation

$$\zeta = 1 - \exp(-\pi vI/\gamma^3 t^{4/3})$$

with  $vI$  the sum of the separate contributions  $vI^H$ ,  $vI^B$ ,  $vI^E$ , etc. The grain size dependence of the half-time will be approximately that characteristic of the type of nucleation making the largest contribution to the total  $vI$ . When saturation occurs, the sites which saturate first must either be those with largest  $vI$  or those of lower "dimensionality"; e.g. if edges have the highest  $vI$ , corners may saturate first, but not boundary surfaces. After one type of site has saturated, only those of higher dimensionality remain; these will eventually control the reaction if their volume nucleation rates are comparable to that of the site which initially had the largest  $vI$ . If one type of site saturates early in the reaction, the sites of higher dimensionality which remain will either contribute little to the overall transformation rate (if they have small  $vI$ ), or else they will also saturate early in the reaction. The half-time will thus be given by  $0.1(L^B/\gamma)$  if any site saturates, and by  $(vI\gamma^3)^{-1/4}$  if none of the sites saturate.

## 58. TRANSFORMATIONS WITH PARABOLIC GROWTH LAWS

The kinetics of a reaction in which the growth rate is controlled by long-range diffusion processes, as in Section 54, will now be considered. Equation (54.9) shows that for a spherical particle, nucleated at time  $\tau$ , the volume at time  $t$  will be given by

$$v_\tau = (4\pi/3)(\alpha_3)^3 D^{3/2}(t-\tau)^{3/2} \quad (t > \tau). \quad (58.1)$$

In most applications, it is assumed that the fraction transformed may be obtained by the method used previously, which gave

$$\zeta = 1 - \exp\left[-\int_0^t v_\tau vI d\tau\right]$$

for a linear growth process in which the nucleation is effectively randomly distributed throughout the assembly. For a diffusion controlled reaction, such as continuous precipitation, however, we can no longer define  $\zeta$  as the fraction of the whole assembly which has transformed. If  $V^\beta$  is the equilibrium volume of  $\beta$  in the whole assembly of volume  $V$ , conservation of solute atoms requires

$$V^\beta/V = (c^m - c^\alpha)/(c^\beta - c^\alpha).$$

During reaction, the total volume of  $\beta$  may be written  $V^\beta(t)$ , and the volume fraction of transformation is now

$$\zeta = V^\beta(t)/V^\beta = \{V^\beta(t)/V\} \{(c^\beta - c^\alpha)/(c^m - c^\alpha)\}. \quad (58.2)$$

In the particular case of a constant number of pre-existing nuclei,  ${}^vN$  per unit volume, we have

$$\zeta = {}^vN v_\tau (c^\beta - c^\alpha)/(c^m - c^\alpha) \quad (58.3)$$

whereas, more generally, the method previously used to treat impingement would give

$$\zeta = 1 - \exp\left[-\{(c^\beta - c^\alpha)/(c^m - c^\alpha)\} \int_0^t v_\tau {}^vI \, d\tau\right]. \quad (58.4)$$

This leads to the usual form of transformation equation, with  $n = \frac{5}{2}$  for  ${}^vI$  constant, and  $n = \frac{3}{2}$  for early site saturation of randomly placed heterogeneous nuclei.

Clearly, however, this treatment of the isothermal reaction rate cannot be justified without much more detailed discussion. As pointed out in Section 54, two growing regions interfere with each other when the volumes from which they are drawing solute atoms begin to overlap. This "soft impingement" is a diffusion problem, and must be treated by solving the diffusion equation with an appropriate boundary condition. Until this has been done, we cannot predict with certainty whether or not eqn. (58.4) will prove to be a reasonable approximation in the early stages of precipitation.

The first approximate treatment of the soft impingement of spherical particles growing parabolically was given by Wert and Zener (1950). The calculation applies only to solutions with a low degree of supersaturation, as defined by (54.14), and assumes moreover that all the nuclei are present at the beginning of transformation. This latter assumption, corresponding to Avrami's model when  $\nu_1$  is very large (p. 20), is not so restrictive as first appears. Since the volume nucleation rate  ${}^vI$  is so sensitive to the degree of supersaturation (Chapter 10), a very small amount of continuous precipitation will change the mean concentration of solute in the untransformed matrix by an amount sufficient to decrease  ${}^vI$  by one or more orders of magnitude. This conclusion is also supported by experimental evidence, since results obtained for continuous precipitation of the type discussed here can generally be interpreted only by assuming that all nuclei were present at the beginning of transformation.

For a solution which is only slightly supersaturated, we use the steady-state approximation (p. 492) for the concentration gradient in the  $\alpha$  phase at the  $\beta$  interface. In order to treat the impingement problem, we must assume that a  $\beta$  particle is growing, not in an

infinite matrix, but in an  $\alpha$  phase where the concentration tends to a steady mean value in regions remote from an interface. For slightly supersaturated solutions, the interfaces will all move sufficiently slowly for the steady-state solution to be appropriate, and the distances between particles of precipitate will all be large, so the assumption seems reasonable. Writing  $c^\infty(t)$  for the mean concentration of  $B$  atoms at large distances from any precipitate at time  $t$ , we have

$$c^\infty(0) = c^m \quad \text{and} \quad c^\infty(\infty) = c^\alpha$$

and eqn. (54.3) becomes

$$(c^\beta - c^\alpha) (dr^I/dr) = D\{c^\infty(t) - c^\alpha\}/r^I. \quad (58.5)$$

Now if all nuclei are present at time  $t = 0$ , and are well separated, the final size of all regions will be the same and may be written  $r^f$ . Then the volume fraction is  $\zeta = (r^I/r^f)^3$ , and we also have the relation  $c^\infty(t) - c^\alpha = (c^m - c^\alpha)(1 - \zeta)$ . Substituting these relations into (58.5) gives

$$\frac{d\zeta}{dt} = \frac{3D}{(r^f)^2} \frac{(c^m - c^\alpha)}{(c^\beta - c^\alpha)} \zeta^{1/3}(1 - \zeta). \quad (58.6)$$

The relation of this equation to our former treatment of impingement may readily be seen. For spherical growth in the absence of impingement, we have

$$\zeta = kt^{3/2} \quad (58.7)$$

if all the nuclei are present at time  $t = 0$ . The corresponding growth rate is

$$d\zeta/dt = \left(\frac{3}{2}\right)kt^{1/2}. \quad (58.8)$$

The usual treatment of the impingement problem is equivalent to multiplying this equation by  $(1 - \zeta)$

$$d\zeta/dt = \left(\frac{3}{2}\right)kt^{1/2}(1 - \zeta) \quad (58.9)$$

and this leads to  $\zeta = 1 - \exp(-kt^{3/2})$ .

Equation (58.8) can also be written in the completely equivalent form

$$d\zeta/dt = \left(\frac{3}{2}\right)k^{2/3}\zeta^{1/3} \quad (58.10)$$

and we might thus, by analogy, expect that impingement can be included by writing

$$d\zeta/dt = \left(\frac{3}{2}\right)k^{2/3}\zeta^{1/3}(1 - \zeta). \quad (58.11)$$

Equation (58.6) is of this form, with

$$k^{2/3} = 2(c^m - c^\alpha) D / (c^\beta - c^\alpha) (r^f)^2. \quad (58.12)$$

Note that whilst (58.8) is equivalent to (58.10), (58.9) is *not* equivalent to (58.11) although they approach each other as  $t \rightarrow 0$ . Wert and Zener plotted  $\zeta$  against  $t$  from (58.6) by nume-

tical integration† and the resulting curve is compared with that given by eqn. (58.9) in Fig. 12.6. There is appreciable deviation towards the end of transformation.

Although, as we have seen, the assumption of zero nucleation rate is probably justified, it is to be expected that nuclei will not be distributed completely at random. This does not

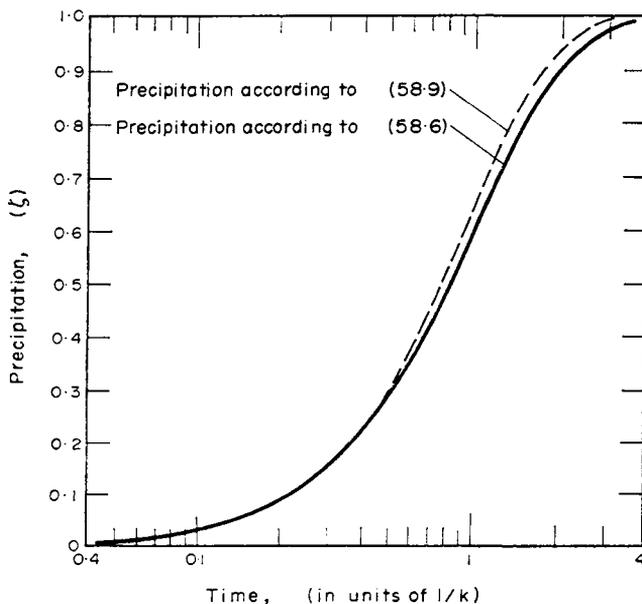


FIG. 12.6. Reaction curve for diffusion controlled growth, eqn. (58.7) compared with the Avrami type curve eqn. (58.9) (after Wert and Zener, 1950).

affect the analysis, since it may be shown that  $c^\infty(t)$  is almost constant despite fluctuations in the density of nuclei. If there are  $vN^C$  nuclei per unit volume, the relation

$$vN^C (4\pi/3) (r^f)^3 (c^\beta - c^\alpha) = c^m - c^\alpha,$$

may be used to give the alternative expression

$$k^{2/3} = (8\pi vN^C r^f D)/3 \tag{58.13}$$

and eliminating  $r^f$  between (58.12) and (58.13)

$$\frac{D^{3/2} vN^C}{k} = \frac{3}{8\sqrt{2}\pi} \left[ \frac{c^\beta - c^\alpha}{c^m - c^\alpha} \right]^{1/2}, \tag{58.14}$$

so that  $vN^C$  may be calculated from the experimental quantities  $D$  and  $k$ .

† Wert and Zener failed to obtain an analytical solution of (58.6), but one was given by Markovitz (1950). His result is

$$\left(\frac{3}{2}\right)k^{2/3}t = \frac{1}{2} \ln(1+x+x^2)/(1-x)^2 + \sqrt{3} \arctan \frac{-\sqrt{3}x}{2+x}, \text{ where } \zeta = x^3.$$

Doremus (1957) has attempted to generalize the Zener–Wert treatment of the impingement problem to non-spherical particles, making use of the asymptotic solutions (54.11), (54.12) instead of the assumption of linear concentration change in the  $\alpha$  phase. When impingement begins, the usual assumption of randomly distributed particles is equivalent to a reduction of the concentration in remote regions of the matrix by a factor  $(1-\zeta)$ . Using the asymptotic values of  $\alpha_j$  valid for small  $\alpha_j$ , and assuming the equilibrium solubility  $c^\alpha$  is so small that it may be neglected, we then have

$$\left(\frac{\partial c}{\partial r}\right)_{r=r^I} = \frac{2\{c^m(1-\zeta)\}^2}{\pi r^I c^\beta}$$

for one-dimensional growth, and

$$\left(\frac{\partial c}{\partial r}\right)_{r=r^I} = \frac{c^m(1-\zeta)}{r^I}$$

for three-dimensional growth.

Substituting back into (54.3), we find the velocity of the interface is

$$\frac{dr^I}{dt} = \frac{2D(c^m)^2(1-\zeta)^2}{\pi(c^\beta)^2 r^I} \quad (58.15)$$

for one-dimensional growth and

$$\frac{dr^I}{dt} = \frac{Dc^m(1-\zeta)}{c^\beta r^I} \quad (58.16)$$

for three-dimensional growth.

If there are  $vN^C$  nuclei per unit volume, all present at the beginning of transformation, eqn. (58.3) gives

$$\zeta = vN^C (c^\beta/c^m)v, \quad (58.17)$$

where  $v$  is the volume of the particle concerned. For the one-dimensional growth of a plate, growing only on its faces, Doremus puts

$$v = 2\pi R^2 r^I \quad (58.18)$$

at a stage in growth when the semi-thickness is  $r^I$ . From (58.15), (58.17) and (58.18),

$$[\zeta/(1-\zeta^2)](d\zeta/dt) = 8\pi DR^4(vN^C)^2,$$

and integrating from 0 to  $t$ , the growth equation is

$$\ln(1-\zeta) + [\zeta/(1-\zeta)] = 8\pi DR^4(vN^C)^2 t. \quad (58.19)$$

Unfortunately, as pointed out on p. 495, the work of Ham (1958) shows that eqn. (58.15) is incorrect if applied to a plate of finite dimensions growing under diffusion controlled conditions, at least if  $R$  is much smaller than the distance between particles. However, we include eqn. (58.19) since it may be appropriate to transformations in which complete edge impingement has occurred at an early stage. The growth conditions would then approximate to one-dimensional growth.

Similarly, for the three-dimensional growth of spheres, eqns. (58.16) and (58.17) give

$$d\zeta/dt = D(4\pi vN^D)^{2/3} (3c^m/c^\beta)^{1/3} \zeta^{1/3}(1-\zeta).$$

This equation is identical with Zener and Wert's expression for spherical growth (58.6) when the substitution  $4\pi vN^C (r^f)^3 c^\beta = 3c^m$  is made, remembering that  $c^\alpha$  is zero in the present approximation.

Doremus also derived growth laws for the three-dimensional diffusion growth of plates and needles of finite initial dimensions. The volume of a disc growing in three dimensions with negligible initial thickness is given by

$$v = 2\pi R^2 r^f + \pi^2 R (r^f)^2 + \frac{4}{3} \pi (r^f)^3, \tag{58.20}$$

where  $R$  is the initial radius and  $r^f$  the semi-thickness, provided the assumption of constant growth velocity in all directions is made (see p. 485). Combination of eqns. (58.16), (58.17), and (58.20) then gives

$$\int_0^z \frac{z \, dz}{a(1-\zeta)} = \pi^2 D v N^C R t, \tag{58.21}$$

where  $a = c^m/(c^\beta v N^C R^3)$  and  $z = \pi r^f/R$ . The relation between  $\zeta$  and  $z$  is

$$a\zeta = 2z + z^2 + (4z^3/3\pi^2). \tag{58.22}$$

Although eqns. (58.21) and (58.22) have a complex solution, a numerical integration is possible. Ham's work shows that the assumption of a spherically symmetrical diffusion field used in deriving eqn. (58.20) is quite wrong, and the only reason for including eqn. (58.21) here is that as an empirical equation it seems to give a rather good fit with the experimental data for the precipitation of carbon from iron. We shall not reproduce the corresponding equation which Doremus derived for the growth of a cylindrical particle of finite initial length but negligible initial thickness.

We have referred to the results obtained by Ham several times in this section and in Section 54. He has given a more rigorous treatment of the diffusion impingement problem, and he found the time dependence of the precipitation rate for an array of spherical  $\beta$  particles arranged on a regular cubic lattice. The method of solution utilizes the symmetry properties of this array, and introduces the boundary condition that the normal component of the solute flux vanishes on the surface of the cubic "cell" surrounding each particle.<sup>†</sup> Apart from an initial transient, of short duration, the result is identical with that given by the Zener-Wert method. Moreover, the growth law is not appreciably affected by a non-uniform distribution of particles.

Ham's work shows, as already emphasized, that a spheroid grows from an initial infinitesimal size with constant eccentricity provided each precipitating atom remains at the

<sup>†</sup> The procedure used has certain formal similarities with the Wigner-Seitz method of calculating electron energy bands in solids. As in the simpler applications of that method, the boundary condition is simplified by replacing the cubic "cell" by a spherical "cell" of the same volume.

point on the surface at which it settles. This result arises from the marked non-uniformity of the gradient  $\partial c/\partial r$  over the interface, the gradient being greatest at regions of high surface curvature. It follows, correspondingly, that the time dependence of precipitation for diffusion controlled growth of plates or needles of constant eccentricity is qualitatively similar to the Zener–Wert result for spheres, the only differences being in the parameters in the formula.

From Fig. 12.6 we see that the usual expression

$$\zeta = 1 - \exp(-kt^n)$$

still provides a reasonable approximation for the growth law in the early stages of the reaction, although Ham emphasizes that a law of this kind has no fundamental significance in diffusion-limited reactions, except as an approximation for small  $t$ . Analysis of experimental results in the form of  $\log \log[1/(1-\zeta)]$  vs.  $\log t$  curves, as described in Section 57, will still give straight lines initially, even for general precipitation reactions. Prior to Ham's work, it was thought that the value of  $n$  gives useful information about the shape of the precipitating particles. Thus the assumption made by Wert and Zener about the growth laws for plates and needles (p. 489) led to  $n = \frac{5}{2}$  and 2 respectively for particles of these shapes. It now appears that  $n = \frac{3}{2}$  for the diffusion-limited growth of spheroids of any shape (plates, needles, spheres), provided all particles were present at  $t = 0$  and had negligible initial dimensions.

It is also of interest to consider the growth laws for precipitate particles in which one or more of the initial dimensions is finite, since this situation may be produced by special nucleation conditions, or by a stage of interface controlled growth. For long cylinders, thickening radially, Ham finds  $n = 1$ ; this result is also given by the Zener–Wert and Doremus treatments. For rods of nearly constant length, or discs of nearly constant radius,  $n$  is again equal to unity, provided that the long dimensions are small compared with the particle separation. If this condition is not satisfied, the initial transient during the establishment of the diffusion field is not negligible in comparison with the remainder of the transformation. Finally, if diffusion-limited growth begins from particles of initial volume more than about one-tenth of the final volume,  $n$  has a value intermediate between 1 and 1.5.

We conclude this section by considering briefly the very complex problem of the kinetics expected when precipitate particles form on dislocation lines. A semi-empirical equation of the Avrami type was proposed by Harper (1951) as an extension of the Cottrell–Bilby formula (eqn. (54.78)). As already noted, the variation of  $c(t)$  with  $t^{2/3}$  is valid only for the early stages of segregation of solute atoms to dislocations, and Harper suggested that at later times the equation

$$c(t)/c^m = 1 - \exp(-kt^{2/3}), \quad (58.23)$$

where  $k$  is specified from (54.78), should be used. Although this equation seems to give good agreement with experiment in some alloys, more detailed analyses show that the reasoning on which it is based is untenable (Ham, 1959; Bullough and Newman, 1959, 1961, 1962, 1972).

The first complete solution of (54.74) was given by Ham (1959) whose treatment assumed the dislocations to be ideal sinks, as in the Cottrell–Bilby model. The boundary conditions he applied were thus  $c = 0$  at  $r = 0$  and  $\mathbf{I} = 0$  at  $r = r_e$  where  $r_e$  is the effective radius from which each dislocation draws solute. As expected, the initial characteristics of his solution correspond to the pure drift flow treatment, but at large times he obtained a simple exponential variation with time. Bullough and Newman (1972) point out that this result applies to various problems for impurity segregation from a finite volume of crystal provided the accumulation of the impurities at the dislocation is expressed by a simple boundary condition; the experimental observation that first-order kinetics are rarely observed in such processes indicates that the boundary conditions are complex. Ham also showed that the angular terms in  $W_i$  and the assumption of a random distribution of defects rather than a regular array have little effect on the predicted kinetics. An important result of his work was that the  $t^{2/3}$  dependence of (54.78) or (54.82) is valid only at a very early stage of the transformation, and his results also show that precipitation should be complete at a much earlier time than is predicted by the Harper equation. Thus the agreement of (58.23) with experiment must be in some sense fortuitous, since the equation has no firm physical basis.

Bullough and Newman (1959, 1962a) and Meisel (1967) investigated a different problem, namely the kinetics of the formation of a Maxwellian atmosphere at the dislocation core using a gas-like boundary condition of constant concentration in the core radius, or alternatively an interaction potential which is bounded in the core region. For weak interactions  $W_i$ , Bullough and Newman (1959) obtained an analytical solution by treating the drift flow as a perturbation of the diffusion flow; strong interactions were treated later by numerical methods. In all cases, the Cottrell and Bilby  $t^{2/3}$  kinetics were found initially, which illustrates that this result depends only on the dominance of the drift flow and is independent of the boundary conditions. However, even with very strong interactions, the number of atoms removed to the core region of the dislocation does not follow the Harper equation, and eventually first-order kinetics are predicted, i.e. (4.11) with  $n = 1$ . It must therefore be concluded that Maxwellian atmosphere formation does not correspond to the physical situation.

The failure of the simple precipitation model of Ham and the atmosphere model led Bullough and Newman (1962b) to consider more complex precipitation models in which there is a finite rate of transfer of solute atoms from the dislocation core into the precipitate particles. For a model in which discrete particles are nucleated along dislocations, they again found that the kinetics did not follow the Harper equation except in the early stages of the process, and at long times the simple exponential behaviour was again found. However in another model of continuous rod like precipitates forming along the dislocation line, an approximation to the Harper type kinetics was found by using a transfer velocity from core to precipitate which decreases during the precipitation process. Physically, this effect was considered to result from the internal pressure generated by the change of volume when the precipitate is formed; this in turn affects the diffusion coefficient near to the core. Other effects of the same type may also give deviations from first order kinetics; for example the field of the precipitate may partially cancel that of the dislocation. These more complex

boundary conditions and their effect on the kinetics cannot be considered further here and reference should be made to review papers by Bullough (1968) and Bullough and Newman (1970).

The exponential growth law summarized in Avrami's equation (4.11) is valid for linear growth under most circumstances, and approximately valid for the early stages of diffusion controlled growth. Table IX summarizes the values of  $n$  which may be obtained in

TABLE IX. VALUES OF  $n$  IN KINETIC LAW  $\zeta = 1 - \exp(-kt^n)$

(a) Polymorphic changes, discontinuous precipitation, eutectoid reactions, interface controlled growth, etc.

Conditions	$n$
Increasing nucleation rate	$> 4$
Constant nucleation rate	4
Decreasing nucleation rate	3-4
Zero nucleation rate (saturation of point sites)	3
Grain edge nucleation after saturation	2
Grain boundary nucleation after saturation	1

(b) Diffusion controlled growth

Conditions	$n$
All shapes growing from small dimensions, increasing nucleation rate	$> 2\frac{1}{2}$
All shapes growing from small dimensions, constant nucleation rate	$2\frac{1}{2}$
All shapes growing from small dimensions, decreasing nucleation rate	$1\frac{1}{2}-2\frac{1}{2}$
All shapes growing from small dimensions, zero nucleation rate	$1\frac{1}{2}$
Growth of particles of appreciable initial volume	$1-1\frac{1}{2}$
Needles and plates of finite long dimensions, small in comparison with their separation	1
Thickening of long cylinders (needles) (e.g. after complete end impingement)	1
Thickening of very large plates (e.g. after complete edge impingement)	$\frac{1}{2}$
Precipitation on dislocations (very early stages)	$\sim \frac{2}{3}$

various experimental situations. The conditions listed are not meant to be exhaustive; for example, the effects of an external surface which may reduce  $n$  by up to 1 for a foil and 2 for a wire have not been included. It is evident from the table that a kinetic investigation which is limited to the establishment of the value of  $n$  most appropriate to the assumed growth law does not, as once assumed, give sufficient information for the growth habit to be deduced.

#### 59. EFFECTS OF TEMPERATURE: NON-ISOTHERMAL TRANSFORMATIONS

If the kinetics of a transformation are found experimentally at a number of different constant temperatures, a complete isothermal transformation diagram may be drawn. This figure, also known as a time-temperature-transformation (*TTT*) diagram, gives the

relation between the temperature (plotted linearly) and the time (plotted logarithmically) for fixed fractional amounts of transformation to be attained. Thus the complete diagram consists of a number of curves of  $T$  against  $\log t_\zeta$ , where  $t_\zeta$  is the time required for the transformation to reach stage  $\zeta$ . Frequently, only two or three such curves are given in a  $TTT$  diagram, measuring the times ( $t_0$ ,  $t_1$ ) for the beginning and end of transformation, and sometimes for 50% transformation ( $t_{1/2}$ ). However, it must be emphasized that  $t_0$  and  $t_1$  cannot really be measured experimentally, and the quantities plotted are something like  $t_{0.05}$  and  $t_{0.95}$ .

When both nucleation and growth rates are temperature dependent, the isothermal transformation rate will not be a simple function of the temperature. In very many reactions on cooling, the nucleation rate is determined mainly by a Boltzmann type equation with an activation energy which decreases more than linearly with temperature (Chapter 10). This gives a rapidly increasing nucleation rate as the undercooling (or supersaturation) increases. The growth rate, in contrast, is controlled by an activation energy which is nearly independent of temperature, and hence the rate decreases as the temperature decreases. These opposing factors give an overall transformation rate which first increases and then decreases again as the temperature falls, leading to the  $C$  curves characteristic of so many  $TTT$  diagrams. At sufficiently low temperatures, the nucleation rate may be so large that the nucleation sites saturate early in the reaction. The growth rate alone then controls the overall reaction rate.

The isothermal reaction curves are usually interpreted in terms of eqn. (4.11), which we have seen is a good approximation in almost all modes of transformation. For most transformations, the value of  $n$  is independent of temperature over appreciable temperature ranges, as is expected from the previous analysis. Since  $n$  depends only on the growth geometry, it should only change when this geometry alters; this happens, for example, when a heterogeneously nucleated reaction reaches the degree of supercooling at which the nucleation sites are saturated at an early stage of transformation.

The value of  $k$ , in contrast, is found to vary markedly with temperature. Differentiation of eqn (4.11) leads to an expression for the rate of transformation

$$d\zeta/dt = nk(1-\zeta)t^{n-1} \quad (59.1)$$

in which  $(1-\zeta)$  is the impingement factor, and  $nk t^{n-1}$  gives the rate law in the absence of impingement. Many writers have used the (rather misleading) analogy of a chemical rate equation, and interpreted the factor  $nk$  of this equation in the same way as a chemical rate constant. A plot of  $\log(nk)$  against  $1/T$  is then made (see Chapter 3) with the object of deriving an activation energy for the reaction.

In the early stages of the transformation, the reaction rate is often assumed to be controlled by two different activation energies, one for formation of critical nuclei ( $\Delta G_c$ ) and the other for their subsequent growth  $\Delta_a g^*$ . The time to some fixed amount of transformation would then be given by

$$\ln t_\zeta = C_{21} + (\Delta G_c + \Delta_a g^*)/kT. \quad (59.2)$$

Note that this is really a form of Becker's equation (49.2), since the nucleation rate and the initial transformation rate are interdependent. The reciprocal rate plot is then

$$\frac{d \ln t_c}{d(1/T)} = \frac{\Delta_a g^*}{k} + \frac{\Delta G_c}{k} + \frac{1}{kT} \frac{\partial \Delta G_c}{\partial (1/T)}. \quad (59.3)$$

A plot of  $\ln t_c$  against  $1/T$  will have a C-shape, as noted above, but will approximate to a straight line at low temperatures when  $\Delta G_c \ll \Delta_a g^*$ . The slope of the straight line may be less than  $\Delta_a g^*$  if  $\Delta G_c$  decreases to zero more rapidly than the (negative) third quantity on the right-hand side of this equation.

Interpreting the initial transformation rate in terms of our equation (4.11), we see that before saturation occurs  $k$  is proportional to the nucleation rate to the power one and to either a growth rate or a diffusion coefficient to the power  $(n-1)$ . Thus we have

$$\frac{d \ln k}{d(1/T)} = \frac{-(n-1)\varepsilon}{k} + \frac{\Delta G_c}{k} + \frac{1}{kT} \frac{\partial \Delta G_c}{\partial (1/T)}, \quad (59.4)$$

where  $\varepsilon$  is the growth or diffusion coefficient activation energy. Note that this equation gives a more accurate indication of the roles played by the two kinds of activation energy in determining the overall transformation rate than does (59.3). In the region where the curve becomes a straight line, nucleation saturation has occurred, and the value of  $n$  has decreased to a lower value  $n'$ . The relation is then

$$\frac{d \ln k}{d(1/T)} = \frac{-n'\varepsilon}{k} \quad (59.5)$$

from which  $\varepsilon$  may be determined.

Equation (59.5) implicitly assumes that the number of nuclei formed is independent of temperature, so that the other factors determining  $k$  (for example eqns. (56.3) or (56.6) are constant. In the same way, for continuous precipitation processes  $k$  depends on the mean diffusion distance between precipitate particles, so eqn. (59.5) should strictly be written

$$\left[ \frac{d \ln k}{d(1/T)} \right]_{c, L^B} = \frac{-n'\varepsilon}{k}$$

to emphasize that the grain size  $L^B$  or the mean diffusion distance should be kept constant. Since these quantities usually have strong temperature variation (because of nucleation rate variation, or grain size variation), results for activation energies frequently have no significance unless special precautions are taken. One way of doing this is to start transforming at a fixed low temperature, to produce nucleation site saturation, and then to measure  $k$  for subsequent transformation at various higher temperatures. This effectively measures the kinetics of the growth part of the transformation only.

Since the pioneer work of Davenport and Bain (1930), *TTT* diagrams have been widely used in industry as a guide to heat treatment procedures. Nevertheless, in industrial practice, the kinetic behaviour of an assembly at constant temperature is frequently of less

importance than its behaviour during constant heating or cooling through a transformation range. The general theory of transformation kinetics is largely confined to isothermal reaction, but there have been several attempts to predict the course of a non-isothermal reaction from an experimentally determined set of isothermal transformation curves. The idealized problem is thus to calculate the  $\zeta-t$  curves from the isothermal  $\zeta-t$  curves and some given  $t-T$  relation.

The difficulties in treating non-isothermal reactions are mainly due to the independent variations of growth and nucleation rate with temperature, mentioned above. The problem is tractable only when the instantaneous transformation rate can be shown to be a function solely of the amount of transformation and the temperature. This leads to the concept of additivity which we shall now describe.

Consider the simplest type of non-isothermal reaction, obtained by combining two isothermal treatments. The assembly is transformed at temperature  $T_1$ , where the kinetic law is  $\zeta = f_1(t)$  for a time  $t_1$ , and is then suddenly transferred to a second temperature  $T_2$ . If the reaction is additive, the course of the transformation at  $T_2$  is exactly the same as if the transformed fraction  $f_1(t_1)$  had all been formed at  $T_2$ . Thus if  $t_2$  is the time taken at  $T_2$  to produce the same amount of transformation as is produced at  $T_1$  in a time  $t_1$ , we have  $f_1(t_1) = f_2(t_2)$ , and the course of the whole reaction is

$$\left. \begin{aligned} \zeta &= f_1(t) && (t < t_1), \\ \zeta &= f_2(t+t_2-t_1) && (t > t_1). \end{aligned} \right\} \quad (59.6)$$

Suppose that  $t_{a1}$  is the time taken to produce a fixed amount of transformation  $\zeta_a$  at  $T_1$ , and  $t_{a2}$  is the corresponding time to produce the same amount of transformation at  $T_2$ . Then in the composite process above, an amount  $\zeta_a$  of transformation will be produced in a time

$$t = t_{a2} - t_2 + t_1 \quad (59.7)$$

if the reaction is additive. The time can also be found from the rule

$$\frac{t_1}{t_{a1}} + \frac{(t-t_1)}{t_{a2}} = 1 \quad (59.8)$$

and we can see that (59.7) and (59.8) are equivalent provided  $t_1/t_2 = t_{a1}/t_{a2}$ . We shall see that when (59.6) is true, this condition is satisfied.

An additive reaction thus implies that the total time to reach a specified stage of transformation is obtained by adding the fractions of the time to reach this stage isothermally until the sum reaches unity. The generalization of eqn. (59.8) to any time-temperature path is clearly

$$\int_0^t \frac{dt}{t_a(T)} = 1, \quad (59.9)$$

where  $t_a(T)$  is the isothermal time to stage  $\zeta_a$  (as plotted on a *TTT* diagram), and  $t$  is the time to  $\zeta_a$  for the non-isothermal reaction. Note that an additive reaction does *not* imply (as

sometimes stated) that for the two-temperature transformation path above  $\zeta = f_1((t_1) + f_2(t - t_1)$ .

It is obvious that (59.6) will be true if the reaction rate depends only on  $\zeta$  and  $T$ , that is, only on the state of the assembly, and not on the thermal path by which it reached that state. An analytical proof of eqn. (59.9) also follows from this statement. Consider a transformation for which the instantaneous reaction rate may be written

$$d\zeta/dt = h(T)/g(\zeta), \quad (59.10)$$

where  $h(T)$ ,  $g(\zeta)$  are respectively functions only of temperature and volume fraction transformed. Then we may write

$$\int h(T) dt = \int g(\zeta) d\zeta = G(\zeta) \quad (59.11)$$

for any transformation path. This equation is equivalent to

$$\zeta = F\left\{\int h(T) dt\right\} \quad (59.12)$$

and in particular for an isothermal transformation

$$\zeta = F\{h(T)t\}. \quad (59.13)$$

According to eqn. (59.13), the fraction transformed at a fixed temperature is dependent only on the time and on a single function of the temperature. This function  $h(T)$  might specify the growth rate or the diffusion coefficient, for example. Transformations at different temperatures then differ only in the time scale, as we assumed above in asserting the identity of (59.7) and (59.8). Reactions of this type are called isokinetic, after Avrami (1939, 1940). Avrami defined an isokinetic reaction by the condition that the nucleation and growth rates are proportional to each other (i.e. they have same temperature variation). Following Cahn (1956b), however, we shall take (59.12) as a more general definition of an isokinetic reaction. We now show that an isokinetic reaction is additive in the sense defined by (59.9).

From (59.11) we see that

$$h(T) = G(\zeta_a)/t_a(T)$$

and on substituting into (59.10)

$$t_a(T) (d\zeta/dt) = G(\zeta_a)/g(\zeta). \quad (59.14)$$

Now consider the integral

$$\int_0^t \frac{dt}{t_a(T)} = \int_0^\zeta \frac{d\zeta}{t_a(T) (d\zeta/dt)} = \frac{G(\zeta)}{G(\zeta_a)}. \quad (59.15)$$

This gives the relation between the time  $t$  and the fraction transformed for the whole non-isothermal reaction, and is the general expression of the concept of additivity. In particular, if  $\zeta = \zeta_a$ , (59.15) reduces to (59.9).

Avrami's condition for an isokinetic reaction will rarely be satisfied, since a fortuitous coincidence is required. For many reactions, however, we have seen that the nucleation rate may saturate early in the transformation. Provided that the growth rate at any instant is then only dependent on the temperature, the reaction will be isokinetic in the general sense defined by Cahn. In the case of discontinuous reactions with duplex cell formation, a direct test is possible if the assembly is allowed partially to transform at one temperature and then quenched suddenly to another temperature and held there for a further time. An abrupt change in the interlamellar spacing indicates that the growth rate  $\dot{Y} = h(T)$ , but a gradual change, as has also been observed, implies that  $\dot{Y}$  depends also on  $dT/dt$ . If this is so, the reaction cannot be additive.

In continuous reactions, the nucleation often saturates at a very early stage. The size of a transformed region is then proportional to a diffusion coefficient, raised to an appropriate power, and to a growth factor  $\alpha$ , which is some function of the concentrations, as described in Section 55. A condition for additivity is thus that this factor is not dependent on temperature. If the compositions are not constant, a change from one temperature to another will not only change  $\zeta$  but disturb the equilibrium at the interface of the already precipitated regions. Thus additivity can only be expected at large degrees of undercooling, where approximate constancy of equilibrium concentrations may be found. This factor is not so important in discontinuous reactions, since the average compositions of transformed and untransformed regions are the same. Strictly, the transformed regions in such reactions will not all be in equilibrium without further diffusional adjustments if the compositions of the final phases vary with temperature.

A more important restriction on the notion of additivity must now be mentioned. A reaction in which the nucleation sites saturate can only be written in the form (59.12) if the number of nuclei is not a function of temperature. Thus for grain boundary nucleated reactions, eqn. (59.9) is valid only if the grain size is constant for all the isothermal transformation data (which give  $t_a(T)$ ) and also for the non-isothermal process considered. The dependence of transformation rate on grain size after saturation for various types of nucleation has been given previously (eqns. (57.2), (57.5), and (57.6)). From these equations we can modify (59.9) for the case where the non-isothermal specimen has a grain size different from that used in the isothermal experiments. For nucleation on grain boundaries, edges or corners, the integral in eqn. (59.9) has to equal the ratio of the non-isothermal to the isothermal mean grain diameter when the transformation reaches stage  $\zeta_a$ . For nucleation on preferred sites randomly distributed through the volume, this integral equals the cube root of the ratio of the number of such sites per unit volume of the isothermal specimens to the number per unit volume of the non-isothermal specimen.

From eqns. (57.2), (57.5), and (57.6) we may also write down expressions for  $G(\zeta)$  by direct comparison with (59.11). These equations are

$$\left. \begin{aligned} G(\zeta) &= -\ln(1-\zeta)/(2^v O^B), \\ G(\zeta) &= [-\ln(1-\zeta)/(\pi^v L^E)]^{1/2}, \\ G(\zeta) &= [-3 \ln(1-\zeta)/(4\pi^v N^C)]^{1/3}, \end{aligned} \right\} \quad (59.16)$$

for nucleation on surfaces, edges and corners of grains respectively. Combining these equa-

tions with (59.15),

$$\left. \begin{aligned} \ln(1-\zeta) &= \ln(1-\zeta_a) \int_0^t [dt/t_a(T)], \\ \ln(1-\zeta) &= \ln(1-\zeta_a) \left\{ \int_0^t [dt/t_a(T)] \right\}^2, \\ \ln(1-\zeta) &= \ln(1-\zeta_a) \left\{ \int_0^t [dt/t_a(T)] \right\}^3, \end{aligned} \right\} \quad (59.17)$$

for the three types of site. This last set of equations gives directly the amount of transformation  $\zeta$  in terms of the experimentally determined quantities  $t_a(T)$  and the path ( $t-T$ ) of the reaction. They are readily modified to allow for a difference of grain size between the non-isothermal and the isothermal specimens.

#### REFERENCES

- AVRAMI, M. (1939) *J. Chem. Phys.* **7**, 1103; (1940) *Ibid.* **8**, 212; (1941) *Ibid.* **9**, 177.  
 BULLOUGH, R. (1968) *The Interactions Between Point Defects and Dislocations*, p. 22, HMSO, London.  
 BULLOUGH, R. and NEWMAN, R. C. (1959) *Proc. R. Soc. A* **249**, 427; (1961) *Phil. Mag.* **6**, 403; (1962a) *Proc. R. Soc. A*, **266**, 198; (1962b) *Ibid.* **266**, 209; (1970) *Rep. Prog. Phys.*, **33**, 101.  
 CAHN, J. W. (1956a) *Acta metall.* **4**, 449; (1956b) *Ibid.* **4**, 572.  
 DAVENPORT, E. S. and BAIN, E. C. (1930) *Trans. Am. Inst. Min. (Metall.) Engrs.* **90**, 117.  
 DOREMUS, R. H. (1957) *Acta metall.* **5**, 393.  
 HAM, F. S. (1958) *J. Phys. Chem. Solids* **6**, 335; (1959) *J. Appl. Phys.* **30**, 915.  
 HARPER, S. (1951) *Phys. Rev.* **83**, 709.  
 HUSBAND, J. N., MASON, K. L., and CHRISTIAN, J. W. (1959) (Unpublished work).  
 JOHNSON, W. A. and MEHL, R. F. (1939) *Trans. Am. Inst. Min. (Metall.) Engrs.* **135**, 416.  
 MARKOVITZ, H. (1950) *J. Appl. Phys.* **21**, 1198.  
 MEISEL, L. V. (1967) *J. Appl. Phys.* **38**, 4780.  
 WERT, C. and ZENER, C. (1950) *J. Appl. Phys.* **21**, 5.

## CHAPTER 13

# *Growth From the Vapour Phase*

### 60. GROWTH OF A PERFECT CRYSTAL AT LOW SUPERSATURATION

The formulation in 1949 of the dislocation theory of the growth of a crystal from the vapour phase or from solution marks one of the major turning points in the development of the modern science of materials. Apart from its intrinsic importance, the theory led to the first direct evidence for the existence of dislocation lines in crystals and it provides one of the most striking examples of the influence which small defect concentrations may have on macroscopic properties.

The dislocation theory was developed after it became evident that there were serious discrepancies between the classical theory and the experimental evidence. For crystals bounded by nearly flat atomic faces, the classical theory of growth was based on the repeated nucleation of successive layers, and the rate of growth was predicted to be negligible unless the supersaturation in the vapour is high, contrary to observation. The classical theory is developed in this section because although it refers to conditions seldom encountered in practice it nevertheless provides the essential background for other growth regimes. The dislocation theory is considered in some detail in Section 61 and the growth of thin films on substrates, which has now assumed very great technological and scientific importance, is discussed in Section 62. Modern evaporation, sputtering and “beam” techniques produce deposition from the vapour at high effective driving forces, so that defects may no longer be necessary for growth. However, the substrate must catalyse the formation of either three-dimensional nuclei of the new solid film (e.g. in the form of spherical caps, as in Section 51) or as closed monolayer regions (e.g. circular “pillbox” shapes) which form effectively two-dimensional step nuclei. In either case, the nuclei may form randomly over the substrate or at specific sites representing defects in the surface structure.

There are three different modes of growth by vapour deposition on substrates. In one such mode, the first atoms or molecules deposited form a complete monolayer on the surface, and the film then grows by completion of successive monolayers. A rival process is so-called “island” growth, in which small two- or three-dimensional clusters are nucleated independently on the substrate and grow into islands of the deposited phase, making contact with each other only at a later stage. The third (or Stranski–Krastanov)

growth mode has intermediate characteristics; the first monolayer, or first few monolayers, is completed before further growth, after which there is a switch to the island mode. These different types of growth are considered in Section 62, but only layer growth is expected for a bulk crystal (i.e. on its own substrate) at low supersaturations. At high supersaturations, classical theory encounters other difficulties which mainly arise from the very small number of atoms in a critical two-dimensional nucleus, and atomistic theories of growth are then required. Finally, some consideration is given in Section 63 to special growth conditions which lead to the production from the vapour phase of, for example, nanocrystalline materials, metallic whiskers or amorphous metallic structures.

With the exception of abnormally high temperatures and pressures at and above the critical point, where the vapour and liquid phases are continuous, the formation of a condensed phase from a vapour is a first-order transformation in terms of the (Ehrenfest) thermodynamic classification, mentioned on pp. 226–227. The condensation of the vapour thus involves the separate stages of nucleation of stable droplets or crystallites and their subsequent growth to larger sizes. If droplets or crystallites are produced in an aggregate, there may also be a further stage of coarsening (either by atom transfer or by agglomeration), but as this is driven by the excess surface energy it is usually considered separately, even though it may occur simultaneously with growth.

The classical description of the formation by means of thermal fluctuations of a very small liquid droplet of a solid crystallite within an homogeneous supersaturated vapour phase was considered in some detail in Chapter 10. The theory has been criticized because many of the assumptions made are of doubtful validity, but it has the merit of giving a clear physical picture of the nucleation process. Experimental tests of the theory are very difficult to devise, especially for metallic vapours, so in recent years computer simulations have been much used in attempts to investigate how well the theory works for particular model assemblies. This is just one example of the increasing use of computers in materials science to obtain numerical solution to complex systems of equations. In Part I, various atomistic calculations to simulate the structures of point, line and surface defects were described and reference will be made to results obtained by computer simulation when discussing individual phase transformations. Hence, it may be useful here to digress briefly from the topic of this chapter in order to give an outline description of the main technique used in such simulations.

A computer simulation is “an experiment performed on a computer in which a model for a real system is tested” (Sutton, 1986). Comparison of the results of the simulation with experimental observations on the real system then enables an assessment to be made of the extent to which the various assumptions and approximations in the model system are adequate descriptions of the real systems. The simulation may also make predictions which are difficult or impossible to verify by experiment; for example, the detailed atomic structure of a particular grain boundary or of a solid–vapour interface. Before such predictions can be accepted, it is necessary to verify that they are not sensitive to simplifying assumptions which may have been made about, for example, the nature of the interatomic forces.

Important materials applications of computer simulation include detailed calculations of image contrast in electron microscopy, and macroscopic (continuum) and microscopic (dislocation or atomistic) treatments of crystal plasticity and fracture. However, we shall be mainly concerned with simulations of the atomic structure or behaviour of crystalline or non-crystalline phases, and of interfaces, surfaces, dislocations and other defects. Atomistic calculations of this kind require the specification of an energy function which enables the total internal energy of any configuration to be calculated from the positions of the individual atoms. For metals it is commonly assumed that the energy may be expressed to a sufficient approximation as the sum of two-body interactions between atomic pairs, in which the contribution of each individual pair to the sum depends only on the distance between the two atoms. (The pair potentials are often also described as “central forces”, but it is also possible, in principle at least, to have many-body central forces.) As discussed in Section 16, simple analytical potentials (e.g. Lennard-Jones or Morse) were often used in the early simulations, but more realistic pseudo- or model potentials are now available for many metals. Nevertheless, it is necessary to keep in mind the severe limitations of this representation of the interatomic forces, especially in applications where the volume is not constant.

There are three main techniques used in atomistic computer simulations, two of which are completely deterministic and one which is used to derive results which depend on averaging over a large number of possible individual “complexions” of an assembly where configurational entropy is important. The first procedure is essentially that already described for grain boundary structure on pp. 352–356 and is sometimes called “molecular statics”. The initial configuration is a block containing a number of atoms judged adequate for the problem under investigation (up to  $10^5$  in some of the more complex simulations), each atom being assigned initial coordinates according to some idealized model of the configuration to be studied (for example, the long-range elastic field of a defect may be used to specify the initial coordinates). This structure is then “relaxed”, i.e. the atomic positions are adjusted so as to minimize the internal energy function. The relaxation may be carried out by successive atom-by-atom adjustments (as described on pp. 352–353) or by a more powerful technique (“lattice statics”) in which the atomic positions are adjusted simultaneously rather than successively. It is necessary to impose either rigid or periodic boundary conditions on the outer surfaces of the block; when periodic conditions can be used, the simulation represents an infinite dimension normal to the boundaries concerned. Molecular statics takes no account of either lattice vibrations or zero point energy; it is essentially a variational procedure leading (ideally) to the equilibrium configuration at 0 K. However, in some cases, the relaxation may lead to a local minimum in the configuration, the energy of which is higher than that of the true equilibrium state. The possible mistaken identification of a metastable configuration with the equilibrium state may be minimized if different initial configurations can be relaxed into the same final state. However, the metastable configurations are also often significant, especially if their energies are not much higher than that of the absolute minimum.

The second principal technique is that of “molecular dynamics”, which includes time and temperature effects and thus is able to deal with problems involving the free energy of

the assembly rather than the internal energy. The particles in a computer block (typically up to about  $10^4$  in number) are given initial velocities as well as positions, and their trajectories are computed by numerical integration of Newton's equations of motion. This gives the new positions and velocities after some time interval  $\Delta t$  which must be appreciably less than the time which characterizes any important atomic process; in condensed phases, for example, the time  $\Delta t$  may be taken as about one-tenth of the time of one atomic vibration, so that  $\Delta t \sim 10^{-14}$  s. The new coordinates and velocities are then used as input data for a further numerical calculation after a second interval  $\Delta t$  and the process is repeated. The computer program thus loops repeatedly to give a new configuration at successive time intervals, enabling kinetic events to be studied. A typical such event is an elementary diffusion process, e.g. a vacancy "jump", whilst a more complex example is the transfer of energy outwards and the creation of interstitial and vacancy defects around an atom given an initial "knock-on" during particle irradiation of a solid. Dynamic simulations also may be used to model crystal growth or phase transformations under certain conditions. Thermodynamic properties of the assembly are calculated as time averages of appropriate functions of the positions and momenta of the particles (for example, the average kinetic energy defines the temperature), but such averages clearly have meaning only if taken over a large number (some hundreds) of individual time steps, and the complete simulation may take several million time steps. During the iteration, the total energy of the particles (kinetic plus potential) remains constant.

Molecular dynamics is less likely than molecular statics to result in a metastable configuration rather than in true equilibrium, but the danger is nonetheless always there simply because there are a large number of alternative configurations of an atomic assembly which are not sampled. Thus, whilst the method includes thermal entropic contributions to free energy and simulates the effects of thermal vibrations, it is unable to represent conditions in which configurational entropy is significant. A Monte Carlo simulation, on the other hand, reverses this difference; it takes no account of thermal vibrations, but considers instead the relative probabilities of a very large number of atomic configurations.

Consider, for example, the average interatomic interaction energy for an assembly of  $N$  atoms in a fixed volume and at a fixed temperature (Sutton, 1986). The trial configurations are generated successively from some arbitrary starting point by random displacements of successive atoms or the interchange of each atom in turn with one of its randomly selected nearest neighbours. The potential energy of each new configuration is then calculated from the assumed interatomic potential and compared with that of its immediate predecessor. If its energy is smaller, the configuration is accepted, but if its energy is larger by an amount  $\Delta W$  a comparison is made of the quantity  $B = \exp(-\Delta W/kT)$  with a random number  $R$  between zero and one. The new, higher-energy configuration is then accepted only if  $B > R$ ; otherwise it is rejected, and the preceding configuration is used again as the starting point for the next change. The energy of each accepted configuration contributes to a sum, which will eventually oscillate about some value which represents the mean internal energy at the temperature considered. Sutton states that if  $N$  is a few

hundred atoms, the number of configurations to give a meaningful average is between  $10^5$  and  $10^7$ .

Monte Carlo methods are often used with an Ising model, various forms of which have already been discussed in Part I. The assembly is divided into cells (which may be atoms, lattice sites or other convenient subdivisions) and each cell may have either of two characteristics (e.g. a magnetic atom may have its spin “up” or “down”, an atomic site may be “occupied” or “vacant”, or may contain either an “A” or a “B” atom, etc.). Such a model is frequently used to simulate various types of phase transformation, including crystal growth from the vapour phase, discussion of which is now resumed.

It is generally agreed that the free energy of a cluster of size  $n$ , and more especially of some critical cluster  $n_c$ , is the most important factor controlling the nucleation rate, and criticism of the classical theory has thus centred around the expression developed for this free energy. The kinetic equations assumed to control the growth and decay of clusters have been questioned less frequently, especially since Langer and Turksi (1973) showed by consideration of all possible fluctuations that growth and shrinkage of embryonic liquid clusters occur mainly by addition or removal of single atoms, even in conditions close to the critical point. In a sense, this amounts to a derivation of the Volmer–Becker–Döring equation (48.8) from first principles.

The most dubious assumption of the classical theory is that macroscopic thermodynamic quantities may be assigned to very small clusters (the so-called “capillarity” approximation), which also carries the implication that the cluster has an internal structure equivalent to that of the bulk product phase, and that the interface is sharp and well-defined. There are at least three ways in which these assumptions may be modified. A basic approach is to make no a priori assumption about the cluster configurations, the growth and shrinkage of which are studied by computer simulation using lattice dynamics or Monte Carlo methods with model interatomic forces. Computer modelling has an opposite disadvantage to that of the classical theory, namely that the number of atoms required to simulate an essentially statistical process such as nucleation may be too large to be handled satisfactorily even in modern computers. A second approach is to adapt the so-called “non-classical” theory of Cahn and Hilliard (1957), developed originally for nucleation outside but close to the spinodal curve in an inhomogeneous binary system exhibiting phase segregation. In a Cahn–Hilliard approach, a liquid embryo may be regarded not as a small droplet but as a local high density region within an inhomogeneous vapour in which the variation of density is continuous. There is then no clearly defined interface and the surface energy of the classical theory is replaced by a density gradient term. The free energy change is thus given by an equation similar to eqn. (42.8), with density replacing composition as the variable. The Cahn–Hilliard theory of precipitation from solution will be further developed in Chapter 16; as Pound (1985) remarks, it does not seem to have been much applied to nucleation from the vapour phase.

The third (and most obvious) way of correcting the capillarity assumption is to retain the formalism of the classical theory by including a dependence of the surface free energy on curvature, i.e. on the cluster radius or the number of atoms in the embryo. The necessity for such a correction was recognized by Gibbs (1878), and some early treatments

of this problem for the apparently simpler case of a liquid droplet forming from a supersaturated vapour were summarized on pp. 422–423. These treatments led to the conclusion that the correction term is negative, i.e. the *effective* surface free energy decreases as the size of the droplet decreases. More recently, attempts have been made to test the validity of the capillarity approximation directly for model systems with idealized atomic forces. Even with very simple interactions, however, direct calculation of the energy of a cluster of  $n$  atoms by molecular dynamics (e.g. Lee *et al.*, 1973) or by lattice gas (Ising model) Monte Carlo techniques (Bonnisent and Mutaftschief, 1974; Binder and Kalos, 1980) cannot be made with sufficient accuracy to distinguish between various formulations of the theory. Martin and his co-workers found an ingenious way of avoiding this difficulty by using a method of calculation based on partly overlapping energy distribution functions, in conjunction with a Monte Carlo technique, to compute accurately the *difference* in free energy between clusters containing  $n$  and  $n+1$  atoms interacting with Lennard-Jones type interatomic forces. They used both a two-dimensional (square) and a three-dimensional (simple cubic) Ising model (see pp. 186, 196) so that their results apply most readily to clustering on a surface (considered further below) or to three-dimensional clustering of solute atoms within an initially random solid solution. Nevertheless, confirmation of the capillarity approximation with this model is a useful indication of its validity in vapour–solid nucleation and other cases.

Martin and his co-workers found that the free energy differences between adjacent clusters containing more than about 10 atoms or molecules are consistent with the use of macroscopic values for the specific bulk and “surface” (“edge” in two dimensions) free energies if a curvature correction to the latter is included. This correction was found to be positive for two dimensions (Jacucci *et al.*, 1983) and negative for three dimensions (Perini *et al.*, 1984ab); it appears in the equation for the free energy as an additional term (independent of  $n$ , but temperature-dependent) in the two-dimensional case and as an edge (or step) energy (proportional to  $n^{1/3}$ ) in the three-dimensional case. A further term in the cubic model, independent of  $n$ , is interpreted as a corner, or kink-in-step, contribution.

Other uncertainties in the classical theory concern the possible inclusion of additional terms of statistical mechanical origin in the expression for the free energy. One such term is the Lothe-Pound or Reiss “correction factor” [see eqns. (47.6) and (47.7) and also Reiss (1977)]; another, used by a different group of authors, arises from a model of a droplet near the critical point due to Fisher (1967). Jacucci *et al.* and Perini *et al.* did not consider a Lothe-Pound correction when comparing their simulations with theoretical equations, but added Fisher’s additional statistical term,  $\tau kT \ln n$ , to eqn. (49.3) for the free energy, as suggested by Eggington *et al.* (1971). Although this term is comparatively small, they claim that their calculations are sufficiently accurate both to demonstrate the necessity for its inclusion and to verify that  $\tau$  is close to its theoretical values of 1.25 for two-dimensional and  $-2/9$  for three-dimensional clusters.

As discussed on pp. 440–441, experimental measurements of the nucleation rate are seldom possible and tests of the theory of homogeneous nucleation have been mainly confined to verification that the rate varies with the supersaturation in the manner shown in Fig. 10.3. Homogeneous nucleation of either a liquid or a solid phase from a

supersaturated vapour phase may be obtained in laboratory experiments of the Wilson cloud chamber type in which a supersaturated vapour is suddenly cooled by expansion. In experiments on water vapour, for example, Madonna *et al.* (1961) found a transition from nucleation of the liquid to nucleation of the solid as the temperature after expansion was reduced. The critical supersaturation ratio at which the nucleation rate becomes very large has now been measured for many inorganic and organic vapour phases by the expansion cloud chamber or supersonic nozzle methods mentioned on p. 441, or by a more recent technique, the diffusion cloud chamber, in which the vapour is subjected to a temperature gradient maintained between two horizontal plates, the lower of which is at the higher temperature. (Note, however, that these results are mainly for the nucleation of liquid droplets and, because of the experimental limitations, no results are available for metallic vapours.) Experimental results have usually been analysed by comparing the theoretical and measured values of the critical supersaturation. The theoretical value is obtained from the macroscopic bulk and surface free energies, together with an assumed value for the critical nucleation rate of (say)  $\sim 1 \text{ mm}^{-3} \text{ s}^{-1}$ ; the calculated value of  $i$  is quite insensitive to the assumed rate. (An equivalent test, discussed in the analysis of the original Volmer and Flood (1934) measurements on p. 441, is to use the measured supersaturation to derive a value for the surface free energy and to compare this with the experimental bulk value.) Good agreement between observed values of the critical supersaturation and theoretical values obtained from the simple classical theory is found for most of the expansion and diffusion cloud chamber measurements (see, e.g., Katz *et al.*, 1976), thus indicating that any curvature corrections to the surface free energy and any statistical corrections to the cluster free energy are very small. This is not unexpected as the critical nucleus size calculated from the experimental supersaturation is  $\sim 50$ -100 atoms or molecules (i.e. critical radii  $\sim 1 \text{ nm}$ ) and clusters as large as this may well have the properties of bulk liquid. As mentioned on p. 441, the agreement of experiment and simple theory for most of the nozzle type measurements is not good, but this is often ascribed to the difficulties of interpretation.

More quantitative analysis is possible only if the actual nucleation rate can be measured for different supersaturations. The first measurements of this kind were made for water vapour by Sharaf and Dobbins (1982) who used a highly instrumented cloud chamber to measure the time dependence of temperature, pressure, supersaturation and, in the later stages, the size and number density of clusters. Their analysis indicated that whilst the experimental value of the critical supersaturation for droplet nucleation ( $i \cong 4.1$  at 375 K) agrees with the theoretical estimate, the theoretical value of the pre-exponential term of eqn. (47.13) is about  $2 \times 10^7$  times larger than the measured value, whilst the theoretical saddle point energy of the critical nucleus is larger than the measured value by a factor of 1.54. This comparison is with respect to the simplest form of the classical theory with a constant (macroscopic) value for the surface free energy. Sharaf and Dobbins criticized the concepts of critical nucleation rate and critical supersaturation as applied to expansion chamber type experiments, as their results indicated that these varied considerably with expansion rate. In discussing their results, Pound (1985) suggested that the two incorrect factors in eqn. (48.14) fortuitously cancel each other to give the expected critical

supersaturation, and it follows that the experimental observation of the expected value of  $i$  cannot be considered as very strong evidence for the correctness of classical nucleation theory. Inclusion of Lothe–Pound or similar statistical corrections tends to increase the theoretical pre-exponential factor (see p. 431), and hence to produce a still larger discrepancy. The reason for the discrepancy is probably the neglect of the curvature correction, but a lower than expected frequency factor and free energy of nucleation also suggest the possibility of some form of heterogeneous nucleation during the experiments.

As noted above, at the supersaturation achievable in cloud chamber type experiments, the critical nucleus size is of the order of 100 atoms, and its radius is thus appreciably larger than that of the two-dimensional nuclei at high vapour–solid deposition rates. The equilibrium shape of a small three-dimensional crystallite is expected to conform to the generalized form of Wulff’s theorem (pp. 155–160) and should thus be determined by the anisotropy of the surface free energy. Very small solid particles, obtained from a metallic smoke, often have the expected Wulff shapes (Uyeda, 1991), but this shape may be modified, especially in face-centred cubic (f.c.c.) crystals, by internal twinning. Ino (1948) suggests, for example, that f.c.c. particles smaller than 10 nm may have an equilibrium icosahedron shape, which is achieved by multiple twinning on {111}, rather than the truncated octahedral shape which gives the lowest free energy if the particle is a true single crystal. In any event, the Wulff thermodynamic criterion is applicable only to nuclei or tiny crystallites which have surface energies comparable with their bulk free energies. The shape of a macroscopic crystal is governed in practice by the growth rates of the various possible faces.

Non-metallic crystals commonly grow from the solution or from the melt in the form of convex polyhedra with well-developed plane faces. Crystal habit, which is the name given to the external form of a crystal, shows many remarkable variations when crystals are grown under different conditions, but well-developed faces are always parallel to close-packed (low index) crystallographic planes; this is the famous Bravais rule. Metallic crystals of this type are seldom obtained except by growth from the vapour. In the freezing of a liquid metal or alloy, it is usual for growth to begin at a number of centres and the non-crystallographic surfaces of interference form the grain boundaries. This is an example of allotriomorphic growth, i.e. growth under conditions such that the shape is governed by external factors. The growth of a crystal nucleus from a melt may, however, be investigated separately, e.g. by pouring out the liquid after partial solidification. It is then usually found that each nucleus grows almost exclusively in the direction of certain crystallographic axes, giving rise to the elongated branched structures known as dendrites. Dendritic growth has already been considered in Section 54, and will be discussed in relation to solidification in the next chapter. In this chapter, the exclusive concern is with normal growth, which may be obtained by very slow cooling from the melt, but is almost always obtained in crystal growth from the vapour phase.

As first pointed out by Gibbs, the addition of atoms to a small growing crystal takes place in an essentially discontinuous manner, as each atom has to attach itself to a site in the crystal. The configuration presented by the crystal to the vapour thus varies

periodically with time, as each successive crystal plane is completed. The free energy of the assembly is a discontinuous function of the size of a perfect crystal.

The structure of a crystal surface may depend strongly, *inter alia*, on its crystallographic orientation, the type of bonding in the crystal and the temperature, and it will also be influenced by accidental factors such as lattice defects and impurity atoms. In Section 18, surfaces were divided into two classes, namely close-packed (low index) surfaces which, at low temperatures, are essentially atomically flat, and non-close-packed (higher index) surfaces which may be modelled as steps and terraces of close-packed surface. Modern investigations of surface structure by low-energy electron diffraction (LEED) and other techniques (see Woodruff, 1986 for a concise review) show that for most clean metallic surfaces this picture is essentially correct, so that the structure of the bulk crystal is simply terminated at a geometrical plane. In some cases, small (< 5–10%) contractions in the spacing of close-packed atomic planes forming the first surface layers have been measured, notably for {110} faces in f.c.c. structures and {100} faces in body-centred cubic (b.c.c.) structures, but otherwise there is no evidence of any significant change in structure at the surface. Similar results have been obtained for compounds such as metallic oxides, but not for silicon and germanium. These semiconducting crystals have been very intensively studied because of their technological importance, and surface “reconstruction” by atomic displacements both parallel and normal to the surface planes has been identified. Such reconstruction is thought to be due to the “dangling” covalent bonds which would be left in a simple unreconstructed surface; the consequent changes in bond angles and lengths cause strain in several atomic layers parallel to the surface and make accurate structure determination very difficult. The surface structure is probably dependent on temperature and possibly also on very low impurity (or dopant) levels, so that several different “surface phases” may be stable under appropriate conditions for a given surface orientation. Some evidence for reconstructed surface structures has also been found in a few metallic surfaces, e.g. {100} in f.c.c. gold and b.c.c. tungsten and molybdenum. In the f.c.c. case, the effect has been attributed to the formation of a surface hexagonal close-packed (h.c.p.) phase, whilst in tungsten the rearrangement may be confined to the topmost atomic layer. The reconstruction is probably strongly influenced by impurities and also does not take place at high temperatures.

In discussing the theory of crystal growth, it will be assumed that there is no surface reconstruction, except for the possibility of multilayer surfaces (surface roughening) which may be formed at relatively high temperatures (see below). This is justifiable for most metallic structures and, in any event, the possible modifications to kinetic theories of growth which may be required by surface reconstruction have not yet been developed in any detail.

The early theory of the deposition of atoms on crystal faces is due largely to Kossel and to Stranski (1928), who made detailed calculations for ionic crystals and estimates for homopolar crystals. Only the latter will be considered, as metals are assumed to approximate more readily to this type of binding. Figure 13.1 shows part of a metallic structure containing a kinked step, as described in Chapter 5. For illustrative purposes only, the simple cubic model of a crystal (Kossel crystal) with nearest- and

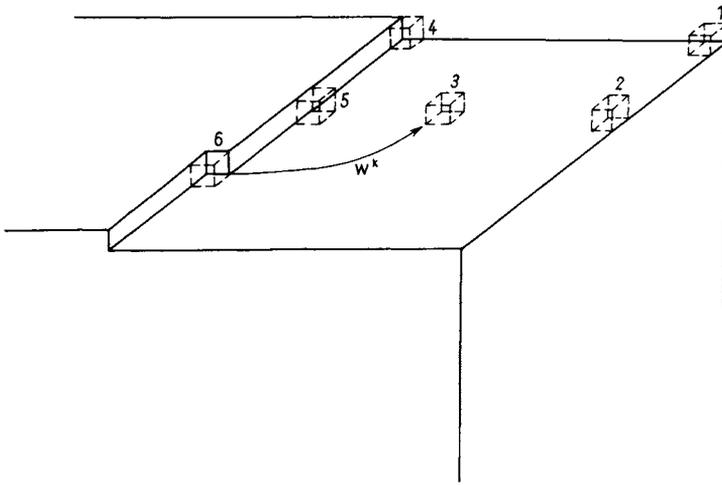


FIG. 13.1. Possible atomic sites on a stepped surface.

next-nearest-neighbour interactions  $2\bar{\epsilon}_1$  and  $2\bar{\epsilon}_2$  will be used. Consider the energy gained by deposition of an atom on any of the sites marked in the figure. At a corner (site 1), this is  $2\bar{\epsilon}_1 + 4\bar{\epsilon}_2$ ; at an edge centre (site 2), it is  $2\bar{\epsilon}_1 + 6\bar{\epsilon}_2$ ; at a face centre (site 3), it is  $2\bar{\epsilon}_1 + 8\bar{\epsilon}_2$ ; at the end of a step (site 4), it is  $4\bar{\epsilon}_1 + 6\bar{\epsilon}_2$ ; at the centre of a step (site 5), it is  $4\bar{\epsilon}_1 + 12\bar{\epsilon}_2$ ; and at a kink in a step, it is  $6\bar{\epsilon}_1 + 12\bar{\epsilon}_2$ . The points at which atoms attach themselves preferentially are thus the kinks in the steps; as stated earlier, the energy required to transfer an atom from such a kink to the vapour phase is the evaporation energy per atom.

A crystal face having steps on it will grow by repeated addition of atoms to the kinks in the steps. This was first appreciated by Kossel (1927), and was named by him the repeatable step ("*wiederholbare Schritt*") for crystal growth. On a perfectly plane crystal face, atoms will tend to attach themselves to positions (3) and to grow together so as to form a raised island surrounded by a step. Such a group of atoms is unstable below a certain size, and hence the growth of a flat crystal face becomes a nucleation problem, the nucleus being two-dimensional.

According to the calculation of Burton and Cabrera (1949), summarized in Chapter 5, a close-packed crystal surface in equilibrium with the vapour is nearly atomically flat at low temperatures and a non-close-packed surface is made up of stepped segments of close-packed surfaces, each step containing a high proportion of kinks. These conclusions are essentially substantiated by the evidence from LEED experiments already mentioned, and also by multilevel computations of the surface structure with model interatomic forces (Weeks and Gilmer, 1976, 1979; van Beijeren, 1977) and by simulations which use Monte Carlo procedures of random interchange of atoms and vacancies to find the minimum energy configurations (Leamy and Jackson, 1971; Leamy *et al.*, 1975). A computer simulation of the configurations and energies of (20,1,0), (10,1,0), (5,1,0) and (100) surfaces of a Kossel crystal (Leamy and Gilmer, 1974) shows that the geometrically necessary steps

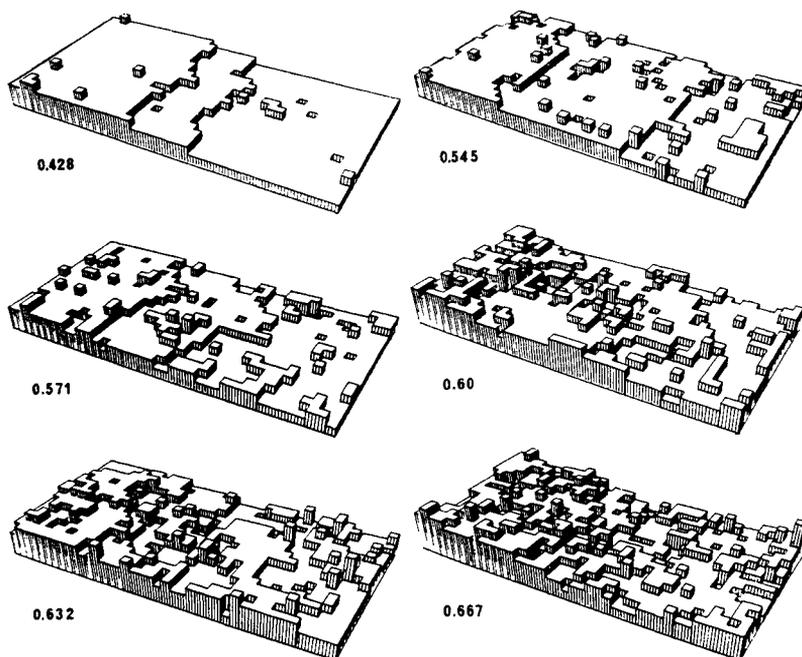


FIG. 13.2. Computer simulation of the structure of a (20,1,0) surface at various values of  $kT/2E_1$  (after Leamy and Gilmer, 1974).

can no longer be distinguished in the general atomic disorder above a temperature  $T_c = (1.28E_1)/k$  which corresponds to the disordering or roughening of the close-packed (100) surface<sup>†</sup> [cf. eqn. (19.1)]. This temperature may alternatively be regarded as that at which the cusp in the Wulff plot disappears (Section 20) and above which surface steps may form without increasing the surface free energy. Some sketches of surface configurations at different values of  $kT/2E_1$  are shown in Fig. 13.2; above the roughening transition, the individual clusters percolate, i.e. they are all interconnected at some level.

As mentioned on p. 152, Burton *et al.* (1949) identified roughening with a form of surface melting, but the tendency in recent work has been to distinguish between the two concepts. Their connection with each other and with models which postulate a greatly enhanced vacancy concentration in the surface atomic layers is discussed by Pontikis and Sindzingre (1987). Roughening involves multilayer occupancy in which long-range correlation of position via lattice translation vectors is nevertheless maintained, whilst surface melting (see p. 154) implies that there is only short-range ordering of atomic positions within a layer. Most of the calculations and simulations which utilize pairwise interatomic potentials indicate that the closest-packed crystal planes remain atomically flat up to the bulk melting temperature, thus confirming the predictions of the simplified

<sup>†</sup>The term "surface roughness" is also used to describe macroscopic variations of surface level (Schey, 1986); these are, of course, unrelated to roughness on an atomic scale.

theory due to Jackson (1958) which was originally developed in order to discuss solid-liquid interfaces. Jackson defined a parameter  $\alpha = \Delta h f / kT$ , where  $\Delta h$  is the change in enthalpy per atom (i.e. the latent heat) for the change from the solid to the fluid phase and  $f$  is the fraction of nearest-neighbour bonds that lie within the close-packed plane under consideration. He then used Bragg-Williams (i.e. mean field) theory to show that atomically smooth surfaces should have  $\alpha > 2$  and rough surfaces should have  $\alpha < 2$ . Hence a surface should not roughen below the melting point if the solid-fluid entropy change per atom exceeds  $2k/f$ , a condition which is satisfied for close-packed planes in solid-vapour equilibrium, but not necessarily in solid-liquid equilibrium.

This conclusion applies only to equilibrium or near-equilibrium conditions; at very high supersaturations, kinetic roughening may occur below the melting point (see Section 62). Experimental attempts to detect a roughening transition by diffraction techniques are very difficult to interpret, but the indirect method of observing the change in morphology from faceted to rounded surfaces (below and above the transition) has been used successfully for several organic or inorganic compounds, such as  $C_2Cl_6$  and  $NH_4Cl$  (Jackson and Miller, 1977), and also for the rare gas crystals krypton and xenon. Although the solid rare gases were grown at low supersaturation, Maruyama (1988) nevertheless found a marked change in morphology at about  $0.8T_m$ , where the close-packed faces  $\{111\}$  and  $\{100\}$  disappeared completely. (There is a further transition to a completely rounded shape just below  $T_m$ , which Maruyama tentatively attributes to surface melting.)

When a stepped face is in equilibrium with the vapour, the rate at which atoms are added to the kinks equals the rate at which they re-evaporate from these sites. In a supersaturated vapour, more atoms reach the steps than leave them and the crystal grows. The growth will change the structure of the interface and such changes can be very important; however, in this section the attention is confined to conditions of small supersaturation where the surface deviates little from its equilibrium structure. Even at low supersaturations, the growth rate of a stepped surface will be quite rapid as there are so many kinks at which atoms may join the steps, and no two-dimensional nucleation is required. This statement is made on the assumption that the concentration of kinks is maintained at or near its equilibrium value despite the atomic flux, and is valid only because the activation energy for kink formation is small (see p. 151). The crystal surface between the steps will have a modified structure which at low vapour supersaturation may be regarded as a simple increase in the number of isolated atoms ("ad-atoms") adsorbed on the surface (see Fig. 13.2). Owing to the low activation energy for surface diffusion, only a small proportion of the atoms added at the kinks will come directly from the vapour. Most of the atoms will be adsorbed on the regions of close-packed surface between the steps, and will then migrate over the surface to the steps. Two processes thus have to be considered: the exchange of ad-atoms with vapour atoms; and the interchange of ad-atoms with atoms in steps through the medium of surface diffusion.

Referring to Fig. 13.1, the fraction of sites occupied by ad-atoms for a crystal face in equilibrium with the vapour will be given by

$$\alpha_0^S = \exp(-w^k/kT) \quad (60.1)$$

where  $w^k$  is the energy required to transfer an atom from a kink to a surface position of type (3). Strictly, the free energy change should be used instead of  $w^k$ , but for atoms and simple molecules where no orientation effects are present the entropy factors may be neglected. For the {100} faces of a Kossel crystal,  $w^k$  is  $4\varepsilon_1 + 4\varepsilon_2$ , and is hence about two-thirds of the evaporation energy  $\Delta h^{sv}$ . For a close-packed face of a f.c.c. crystal, simple considerations show that  $w^k$  is  $6\varepsilon_1$ ; i.e. one-half of the evaporation energy. With a typical value of  $\varepsilon_1 \cong 10.1$  eV,  $\alpha_0^S \cong 10^{-3} - 10^{-4}$  at 600 K. We let  $w^s$  be the energy of evaporation of an ad-atom to the vapour, so that  $\Delta h^{sv} = w^k + w^s$ . The remaining important energy parameter is the activation energy for surface diffusion,  $\varepsilon^S$ , which may be as low as one-twentieth of the evaporation energy.

An ad-atom will tend both to re-evaporate and to diffuse over the surface. By analogy with eqn. (40.8), the surface diffusion coefficient may be written

$$D^S = \frac{\Gamma r_1^2}{4} = (z^S r_1^2 \nu / 4) \exp(-\varepsilon^S / kT) \cong r_1^2 \nu \exp(-\varepsilon^S / kT) \quad (60.2)$$

where  $z^S$  is the surface coordination number,  $r_1$  is the surface nearest-neighbour distance and it has been assumed that the probability of any surface site being occupied by an ad-atom is very much less than unity. The mean time of stay of an ad-atom on the surface will be given by

$$1/\tau^S = \nu \exp(-w^S / kT). \quad (60.3)$$

The frequency factors in these two equations are not necessarily the same but in the absence of marked entropy effects they are both of the order of the atomic frequency of vibration and it will therefore be assumed that they may be equated.

The mean displacement of an atom before re-evaporation, a measure of which is  $x^S = (D^S \tau^S)^{1/2}$  (see p. 389), is thus

$$x^S = r_1 \exp\{(w^S - \varepsilon^S) / 2kT\}. \quad (60.4)$$

From the relative magnitude of  $w^S$  and  $\varepsilon^S$ , it follows that  $x^S$  is much larger than  $r_1$ , and hence surface diffusion is important for the growth of a step. In Chapter 5, eqn. (18.6) gives an expression for the mean distance  $x_\theta$  between kinks in a step, and the rate of advance of a straight step in contact with the supersaturated vapour for the simple case of  $x^S \gg x_\theta$ , a condition which is nearly always satisfied. When the vapour is in equilibrium with the stepped face, the number of atoms evaporating from a given site in unit time will be equal to the product of the fraction of sites occupied ( $\alpha_0^S$ ) and the frequency with which each such atom evaporates ( $1/\tau^S$ ). For the supersaturated vapour, therefore, the number of atoms striking any lattice site in unit time may be expressed either as  $oR$ , where  $o$  is the area of an ad-atom and  $R$  is the incident flux of atoms given by the collision factor  $p/(2\pi mkT)^{1/2}$  from the kinetic theory of gases, or as  $i\alpha_0^S/\tau^S = i\nu \exp(-\Delta h^{sv}/kT)$ , where  $i$  is the supersaturation ratio. On average, all atoms adsorbed within a distance  $x^S$  of a step will reach this step before re-evaporating, and the number of atoms reaching any step in unit time will thus be  $2x^S i\alpha_0^S / r_1 \tau^S$ , the factor of two arising because atoms may approach the step from either side. As the kink density in metals is so high, it may be assumed that

effectively all the atoms which reach the step will be adsorbed on it. For generality in applying the theory to complex molecules involved in the growth of some non-metallic crystals, however, a coefficient  $\beta$  representing the probability of adsorption is introduced.

The number of atoms leaving unit length of a step in unit time is  $2x^S\alpha_0^S/r_1\tau^S$  (i.e. the number absorbed when the vapour is in equilibrium with the crystal), so that the net rate of condensation on the step is  $\beta 2x^S(i-1)\alpha_0^S/r_1\tau^S$ . The velocity with which the step moves forward is thus given by

$$u_\infty = 2(i-1)\beta x^S\alpha_0^S/\tau^S = 2\beta jx^S v \exp(-\Delta h^{sv}/kT) \quad (60.5)$$

substituting for  $\alpha_0^S$  and  $\tau^S$  from eqns. (60.1) and (60.3).

The derivation of this equation is only approximate, particularly in the assumption that all atoms striking the surface within a distance  $x^S$  of a step are captured by that step. The same equation is obtained more rigorously by solving the surface diffusion equation on the assumption that the surface supersaturation ratio  $\alpha^S/\alpha_0^S$  is unity near the steps and  $i$  in regions remote from the steps. Harris (1989) pointed out that Fick's Law is strictly invalid in the presence of adsorption and desorption, and he gave a more microscopic description based on the Fokker-Planck equation (see p. 449). However, his conclusion is that Fick's Law will apply if  $D^S$  is multiplied by a factor  $(1+\zeta)$ , where  $\zeta$  is the ratio of the relaxation time for diffusion to  $\tau^S$  and is usually much less than unity.

Equation (60.5) implies that an isolated straight step will advance perpendicularly to its length with a velocity proportional to the degree of supersaturation in the vapour,  $j$ . If  $D^S$  [eqn. (60.2)] and hence  $x^S$  are independent of orientation, this velocity will not depend on orientation. The assumption that all atoms within  $x^S$  of a step are captured will clearly not be valid for a surface containing a parallel sequence of steps, unless these steps are separated by distances  $y > 2x^S$ . Steps which are relatively close together will compete for the available atoms, and the solution to the surface diffusion equation (again assuming  $x^S \gg x_{ii}$ ) then leads to a corrected expression for the velocity of each straight step,

$$u_\infty = 2\beta jx^S v \exp(-\Delta h^{sv}/kT) \tanh(y/2x^S). \quad (60.6)$$

The growth rate of the face in the direction normal to the close-packed plane (i.e. in the  $x_2$  direction of Fig. 13.3) is given by

$$Y = qd = Y_m(2\beta x^S/y) \tanh(y/2x^S) \quad (60.7)$$

where  $q = u_\infty/y$  is the step flux,  $d$  is the step height and  $Y_m$  is the maximum growth rate and is obtained from the difference in the rate of arrival of vapour atoms and the rate of evaporation as

$$Y_m = jdoR/i = jdv \exp(-\Delta h^{sv}/kT). \quad (60.8)$$

Equation (60.6) assumes that the steps are equally spaced during growth, but it is clear that the possibility of non-uniform spacing must be considered. The velocity of a step depends on the proximity of its neighbours, and under certain growth conditions a uniform distribution of steps may be unstable against local fluctuations in density and velocity, leading to a configuration in which the steps are bunched together. This

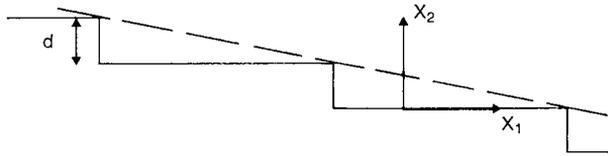


FIG. 13.3. Axes used in theory of step growth.

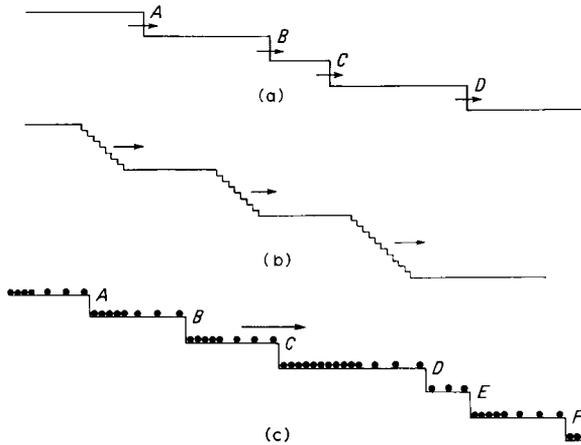


FIG. 13.4. The formation of a platelet structure (after Elbaum and Chalmers, 1955; Frank, 1958): (a) individual atomic steps; (b) step clusters; (c) possible role of absorbed impurities in stabilizing clusters.

possibility was first discussed by Elbaum and Chalmers (1955) in connection with observations of the structure of a decanted liquid–solid interface. Their proposal is illustrated in Fig. 13.4(a). Consider a group of four steps in which by chance *B* and *C* are close together, and so are moving more slowly than the other steps. This means that *D* will move further away from *C*, whilst *A* will catch up with *B*, giving a group of three steps which will move still more slowly. According to Elbaum and Chalmers, this process will continue with more and more fast-moving steps arriving at the rear of the group, until virtually all fast-moving isolated steps have joined a cluster. The distribution will thus tend towards the general form shown in Fig. 13.4(b), in which the aggregates of individual steps are visible under a microscope as large surface steps.

This theory is incorrect if applied to the growth of a pure crystal, and a fuller treatment leads to the result that the clustered configuration is unstable and a regular sequence of equally spaced steps is weakly stable against local fluctuations. The instability of the cluster arises because the step at the head of a bunch, having flat crystal ahead of it, will be able to move faster than the other steps in the bunch, so that it will escape. The composition of the bunch is thus continually changing as it loses steps at its head and gains new steps at its rear. With certain assumptions, it may be shown that an initial cluster of steps will gradually disperse as the individual steps move through it, but that special

conditions may lead to stable clusters. It is also clear that in general the cluster will move with a different velocity from that of the individual steps, in much the same way as the group velocity of a wave packet differs from the wave velocity. The flow of steps across the crystal face is analogous to the movement of road traffic on crowded roads, or of water in flooded rivers, in both of which the flow rate (measured in cars or gallons per minute) may be assumed to depend on the linear density (measured in cars or gallons per mile).

These analogies have more than illustrative value, as a mathematical treatment of the flow of cars or water was developed in advance of the application to crystal growth (Lighthill and Whitham, 1955). The relevance of the calculations was realized independently by Cabrera and Vermilyea (1958) and by Frank (1958), and both sets of authors used the method to discuss crystal growth and dissolution from solution. The basic assumption of these continuum theories is that the step flux depends *only* on the density of steps in the immediate vicinity of the point considered. Following Frank, it is convenient to choose axes  $x_1$  in the close-packed planes normal to the steps and  $x_2$  normal to the close-packed planes. Let  $k^S$  be the number of steps per unit length in the  $x_1$  direction at any point (the step density) and  $q$  the number of steps passing the point in unit time (the step flux). The mean surface slope is

$$\partial x_2 / \partial x_1 = -k^S d \quad (60.9)$$

and the growth rate in the  $x_2$  direction (NB not in the direction normal to the surface) is

$$\gamma = \partial x_2 / \partial t = qd. \quad (60.9a)$$

The velocity of an individual step is a function only of  $k^S$  and may be written

$$u_{\sim}(k^S) = q/k^S. \quad (60.10)$$

From the condition that the total number of steps must be conserved (the continuity equation)

$$\partial q / \partial x_1 + \partial k^S / \partial t = 0$$

and, as  $q$  is assumed to depend only on  $k^S$ , this may be written in the equivalent forms

$$\left. \begin{aligned} c(k^S)(\partial k^S / \partial x_1) + \partial k^S / \partial t &= 0, \\ c(k^S)(\partial q / \partial x_1) + \partial q / \partial t &= 0 \end{aligned} \right\} \quad (60.11)$$

where

$$c(k^S) = dq/dk^S. \quad (60.12)$$

Equations (60.11) and (60.12) show that regions on the surface with constant step density  $k^S$  (and hence with constant flux  $q$ ) move with a velocity  $c(k^S) = dq/dk^S$ , called the kinematical wave velocity in the original paper of Lighthill and Whitham. If the step density  $k^S$  is plotted as a function of the coordinate  $x_1$  and the time  $t$ , it will be constant along a straight line of slope

$$dx_1/dt = c(k^S). \quad (60.13)$$

A line of this type is called by Frank a "characteristic".

The relation between  $x_2$  and  $x_1$  defines the profile of the crystal surface at any time  $t$ , and in the three-dimensional space  $x_1, x_2, t$  this profile generates a surface. The characteristics are lines on this surface, having straight line projections on the  $(x_1-t)$  planes. The projection on the  $(x_1-x_2)$  planes will give the path of a constant density region in ordinary space, and is

$$dx_2/dx_1 = (\partial x_2/\partial x_1)_t + (\partial x_2/\partial t)/(dx_1/dt) = -d(k^S - q/c) \quad (60.14)$$

which is constant because  $k^S$  and  $q$  are constant along a characteristic. The characteristics must thus be straight lines on the  $(x_1-x_2-t)$  surface. Frank points out that eqn. (60.14) shows that, for crystals growing or shrinking with curved surfaces, points of constant orientation have straight line trajectories.

The continuum theory uses the smeared parameters  $k^S$  and  $q$  to represent the behaviour of groups of steps, and is thus incapable of dealing with changes in step distribution which depend on the differences in the behaviour of adjacent steps. An atomic kinematical theory of step growth was developed by Mullins and Hirth (1963) on the assumption that the velocity of any given step depends only on its distances from the steps immediately adjacent to it, and is obtained by adding together the two values of some function evaluated for these two distances. Thus the velocity of the  $i$ th step in a chain is given by

$$u_i = u'(y_{i-1}) + u'(y_i) \quad (60.15)$$

where  $y_{i-1}$  and  $y_i$  are the distances from the adjacent steps. For a clean surface with step velocity limited by surface diffusion, the function  $u'(y)$  of eqn. (60.15) corresponds to  $1/2u_\infty$  of eqn. (60.6).

Some consequences of eqn. (60.15) are immediately apparent. Two isolated steps with initial spacing  $y$  will preserve this spacing and move with constant velocity  $u'(y) + u'(\infty)$ , but an isolated step behind the pair will overtake it and displace the leading step which escapes. The second and third steps then form a pair with a final spacing which was shown by Mullins and Hirth to tend exponentially to  $y$  as the initial distance of the third step from the pair increases. The leading step will always escape from a group of three steps but, if the initial spacings  $y_1$  and  $y_2$  are both much smaller than  $x^S$ , the other two steps will coalesce to form a double step. In a similar way, the leading pair of a train of four steps escapes and the trailing pair either has a constant final separation or collapses into a double step.

Generalization of these results to clusters containing large numbers of steps shows a continued dependence on whether the cluster contains an odd or an even number of steps. When the total number of steps is odd, the approach of the trailing step to the others will always displace the leading step to  $+\infty$ . A finite train of evenly spaced steps will begin to break up by pairwise grouping at the trailing end and by spreading from the front. The escape of the leading step from an even train is impossible unless there is some coalescence of other steps. These results of the Mullins and Hirth treatment modify the conclusions of the continuum theory, which is now utilized to discuss the problem of bunching of large numbers of steps.

In Frank's theory a macroscopic step is represented by a local region of high density  $k^S$ , and the sharpness of the edges of the step depends on the rate of change of  $k^S$  with  $x_1$ . In the limit for a very sharp step, adjacent regions have different densities  $k_1^S$  and  $k_2^S$ , giving a discontinuity of slope according to eqn. (60.9). The path of this discontinuity in the  $(x_1 - t)$  plane will be given by

$$dx_1/dt = (q_2 - q_1)/(k_2^S - k_1^S) \quad (60.16)$$

and this is not necessarily a straight line because the density difference across the discontinuity need not be constant.

Although the detailed relation between  $q$  and  $k^S$  is not known, it follows from eqn. (60.6) that the velocity of an individual step decreases as  $k^S$  increases, so that  $q$  rises less than linearly with  $k^S$  (Fig. 13.5). This means that  $c(k^S)$  also decreases with  $k^S$ , and hence from eqn. (60.13)  $dt/dx_1$  increases with increasing  $k^S$ . Following Frank, consider now a surface with a uniform step density except for a single cluster. If there is a symmetrical variation of  $k^S$  with  $x_1$  at time  $t_0$ , as shown in Fig. 13.6(a), the surface profile [Fig. 13.6(b)] will contain a rounded step. The variation of  $k^S$  with  $x_1$  at some later time  $t_1$  may now be found by drawing the characteristics through various points, as shown in Fig. 13.6(c). These will be parallel straight lines except in the region of the cluster where the slope  $dt/dx_1$  will first increase and then decrease again. The effect of this is shown in Fig. 13.6(c); at the leading edge of the cluster, the characteristics diverge, whilst at the trailing edge they converge and meet. Whenever two characteristics meet, the intervening values of step density have been eliminated, so that a discontinuity of density develops at the trailing edge and moves according to eqn. (60.16). The profile of the step at time  $t_1$  is shown in

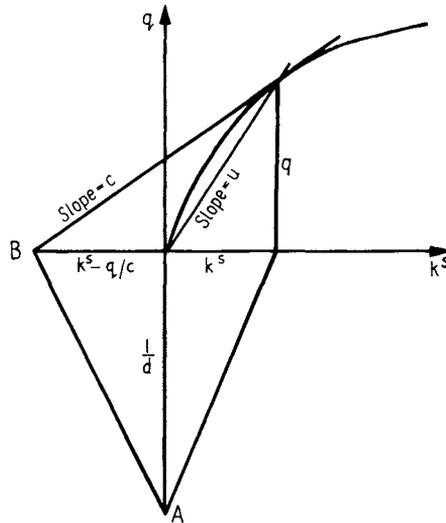


FIG. 13.5. Form of relation between  $q$  and  $k^S$  (after Frank, 1958). The projection of the characteristic in the  $x_1, x_2$  plane (the projection trajectory) is given by  $AB$  [see eqn. (60.14)].

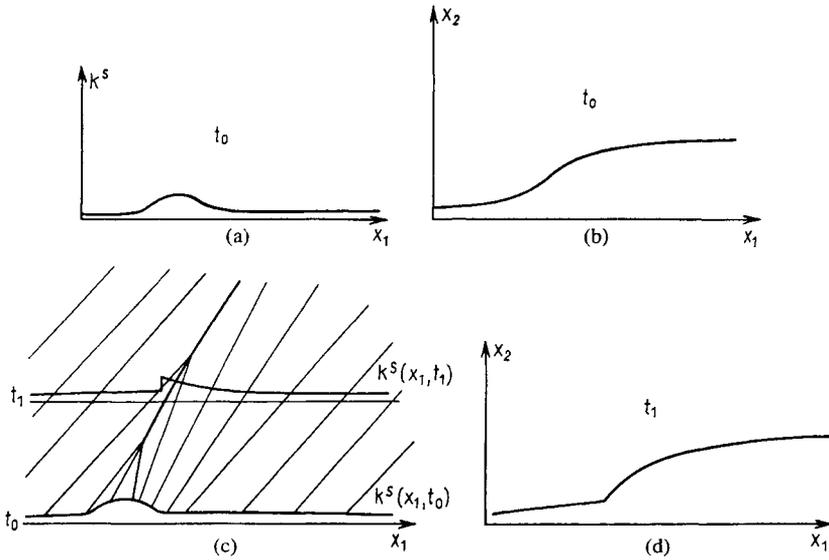


FIG. 13.6. The development of a surface profile (after Frank, 1958): (a) variation of  $k^S$  with  $x_1$  for an isolated cluster; (b) corresponding surface profile; (c) converging characteristics lead to a discontinuity in  $k^S$  at the trailing edge of the cluster; (d) surface profile at a later time  $t_1$ .

Fig. 13.6(d); a discontinuity of slope at the rear is associated with an increasingly diffuse leading edge. Because every value of  $k^S$  at  $t_1$  is derived from the same value of  $k^S$  at time  $t_0$ , it follows that the slope of the profile at  $t_1$ , although it now has a discontinuity, is everywhere between the range of values present initially. As the crystal grows, it simultaneously gains new steps at the rear and loses steps at the front, but the bunch gradually fades out rather than increases.

The theories of Frank, and Mullins and Hirth both lead to the conclusion that the average step spacing in a cluster will increase with time, but that this spacing will decrease at the trailing edge. The atomistic theory predicts that bunching into pairs occurs within the cluster, so that the increase in average spacing represents an increase in the interpair (or even) spacings  $y_{\text{even}}$ . Finally, Mullins and Hirth point out that, at the rear of the cluster, the velocities and orientation trajectories depend on the explicit odd and even spacings,  $y_{\text{odd}}$  and  $y_{\text{even}}$ , and not just on  $1/2(y_{\text{odd}} + y_{\text{even}})$  as in the continuum theory.

A step pile-up or a stable cluster analogous to that postulated by Elbaum and Chalmers is predicted by both theories only if a discontinuity in slope develops at the head of the step instead of the rear. This is not possible for the form of  $(q - k^S)$  relation shown in Fig. 13.5 [i.e.  $d^2q/d(k^S)^2$  negative], but Frank showed how the presence of impurities could modify the growth conditions.

Consider an impurity which is adsorbed on the surface, but is present in small amount so that it takes a relatively long time to reach equilibrium. The impurity is supposed to impede the movement of a step, but is incorporated into the crystal as the step passes. The impedance experienced by any particular step will then depend on the time elapsed since

the step ahead of it passed the same region, i.e. on  $l/q$ . This kind of effect will thus impede steps which are far apart more than steps which are close together, which is opposite to the previously considered influence of the diffusion field of each step. If the effect is sufficiently large, the sign of  $d^2q/d(k^S)^2$  is reversed, and the discontinuity in slope then appears at the head of a bunch as  $q$  increases more than linearly with  $k^S$ .

Applying this theory to the problem of the stability of a uniformly spaced sequence of steps [Fig. 13.3(c)], it now follows that if one step temporarily lags a little behind its proper position, the surface ahead of it will adsorb more impurity, so that its motion will become still slower. The step ahead of it will continue to move at the same speed, or may even accelerate as it has less competition from the diffusion field of the slow-moving step, so that it will move further away. The steps behind the slow-moving one will continue to close up on it to smaller-than-normal distances, as they are meeting clean surface, and will only slow down to the limiting velocity of the slow step when the diffusion field effect forces them to do so. Thus the stability of the uniform distribution has been destroyed, and replaced by a stable clustered distribution, in which the distance between clusters is sufficiently large for adsorption equilibrium to be fully developed before the arrival of each cluster. In the limit, the cluster velocity will be the same as the step velocity, and the composition of each cluster will stay constant. The traffic analogy may again be useful. If all cars can move at the same speed, except when restricted by cars ahead of them, local traffic jams will be unstable and will gradually disappear. A stable series of traffic jams will develop, however, if there are a number of slower-moving vehicles, and each cluster will then move at the speed of the slow vehicle at its head. The stability disappears when overtaking is allowed, but there is no analogy to this in the crystal growth problem; as emphasized by Mullins and Hirth, no single steps can move through a stable cluster and the average step spacing decreases with time.

From the above results it may be concluded that a stepped crystal face will grow or shrink rapidly even at low degrees of supersaturation or undersaturation respectively. There is, however, a significant asymmetry in the processes of growth and evaporation which arises from the fact that formation of a new step at the edge of a flat crystal plane will reduce the free energy during evaporation but not during growth (see Fig. 13.1). This means that flat surfaces are readily converted into stepped surfaces during evaporation and there is no shortage of steps even in defect-free crystals. Hirth and Pound (1957a) supposed that a train of steps originating at an edge during evaporation will accelerate and increase their spacing at a decreasing rate until a steady-state spacing  $y_{ss}$  is attained. This steady-state velocity will be given by eqn. (60.5), where  $j$  has changed sign and now represents the undersaturation, and it will be achieved if  $\tanh(y_{ss}/2x^S) \cong 1$ . Hirth and Pound assumed (rather arbitrarily) that this condition corresponds to  $y_{ss} \cong 6x^S$ , and they used this to deduce a limiting value for the evaporation coefficient which measures the ratio of the actual evaporation rate to the "ideal" rate obtained from the difference of gas collision rates with the surface at the equilibrium and actual vapour pressures. According to the diffusion theory developed above, this evaporation coefficient is given in the limit of free evaporation (zero vapour pressure) by  $(2x^S/y_{ss})\tanh(y_{ss}/2x^S) \cong 3^{-1}$ . In a later paper, Hirth and Pound (1957b) showed that the steady-state value of the evaporation coefficient

should be little changed even for steps which are nucleated from screw dislocations (see Section 61), except at high dislocation densities and low temperatures. However, a basic difficulty of their theory seems to be that although the limiting step velocity is not very sensitive to the final value of  $y_{ss}$ , the evaporation coefficient is nearly linearly dependent on  $y_{ss}^{-1}$ .

Measurements of the evaporation coefficient, e.g. Winterbottom (1967) for silver and Mar and Searcy (1970) for zinc, usually give values close to unity, and thus do not agree with the Hirth-Pound prediction. Surek (1972) used a computer simulation technique to treat the problems of step nucleation and step growth together. A new step is assumed to form from an edge-type source whenever the preceding step has moved a distance  $y_1$  (treated as a parameter of the model), the step interactions are described by eqns. (60.15) and (60.6) and the steps disappear at a sink at a fixed distance from the source. Transient and steady-state solutions for the evaporation rate and the surface topography are thus obtained, and these show that the value of 0.33 for the evaporation coefficient is a very special result, obtained only by making the source emit steps at intervals  $y_1 \geq 5x^S$ . For more rapid nucleation (smaller  $y_1$ ), the fraction of the surface with spacings greater than  $5x^S$  never exceeds 10%, even for a semi-infinite crystal, so that the limiting value of the evaporation coefficient is much larger than 0.33; this effect is naturally accentuated for finite distance between source and sink. On the other hand, as we have indicated above, the evaporation coefficient continues to fall below 0.33 with increasing  $y_1$  because  $y_{ss} > y_1$ .

Small values of the evaporation coefficient can thus only arise from large initial spacings of the steps emitted by a source, or during the transient period before steady-state behaviour is attained. The timescale of this transient stage increases for low source densities, and also depends on  $y_1$ . For  $y_1 > 6x^S$  the steady state is reached by the time the first step reaches the sink, after which the surface profile is invariant with time; however, with smaller values of  $y_1$ , the initial rapid and linear increase of effective evaporation coefficient with time is succeeded by an asymptotic approach to the steady-state value. The steady state is attained first near the source, so that constant nucleation rate is obtained in advance of the unchanging surface profile which is characteristic of the final state.

The shapes of crystals growing, evaporating or dissolving from given initial conditions may be treated in an elegant manner using the methods of the continuum kinematic theory of step growth, and applications include the prediction of transitional shapes and the contours of etch pits. Experimental tests of this theory have been made in a few cases (Frank and Ives, 1960; Ives, 1961; Hullelt and Young, 1966), and it has also been verified by computer simulation (Hullelt and Young, 1966; Surek, 1972) that the atomistic theory of Mullins and Hirth gives predictions in qualitative agreement with those of the continuum theory.

We now return to the theory of crystal growth, in which it cannot be assumed that new steps form readily at crystal edges. An important consequence of the rapid growth of stepped faces is that they tend to disappear from the crystal, leaving only the relatively slowly growing faces. The external form of a perfect crystal is governed by this growth condition, rather than by the minimum surface energy condition mentioned on p. 155.

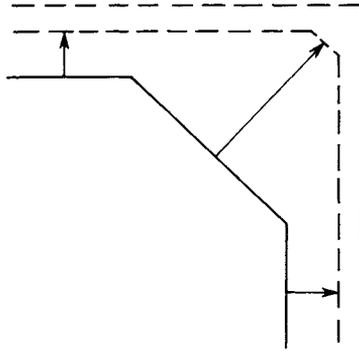


FIG. 13.7. Diagram to show that rapidly growing faces disappear from the crystal form.

Whatever the initial arrangement of crystal faces, the crystal will be bounded eventually by close-packed faces only. This may be seen from Fig. 13.7, which shows one rapidly growing face between two slowly growing faces; the difference in growth rate results in the progressive elimination of the rapidly growing face. The same conclusion is reached by considering the atomic growth process. Each step on the non-close-packed surface is removed by the growth process to the edge of the plane where it disappears. Repetition of the process removes all the steps from the high index plane, and replaces it by the principal (close-packed) planes parallel and perpendicular to the steps.

The ability of high index planes to grow at low supersaturations may thus be ineffective in causing continuous crystal growth as these planes soon disappear and leave only close-packed planes. Thus it is now necessary to consider the problem of crystal growth on essentially flat atomic planes.

Suppose that there are  $n$  ad-atoms or molecules on the surface of a close-packed plane forming the two-dimensional nucleus of a second plane. In Chapter 10 it was noted that a three-dimensional crystal of finite size has a higher equilibrium vapour pressure than an infinite crystal because of the influence of the surface energy. In the same way, the edge energy of a two-dimensional nucleus implies that it is in equilibrium with a higher pressure of vapour than is an infinite plane. If the vapour is supersaturated with respect to the crystal as a whole, the change in free energy consequent on the formation of such a nucleus will pass through a maximum value, as in the analogous three-dimensional case. If the edge free energy per unit length,  $\gamma$ , is independent of direction, the nucleus will be circular and the change in free energy resulting from its formation will be

$$\Delta G' = n(g^s - g^v) + 2\pi r\gamma \quad (60.17)$$

where  $g^s$ ,  $g^v$  are the free energies per atom in the crystal and vapour respectively and  $r$  is the radius of the nucleus. This equation does not include a Lothe-Pound correction factor, so by analogy with the treatment of three-dimensional nucleation in Chapter 10, the free energy change has been written as  $\Delta G'$  and it is supposed that this is related to the total

free energy change on forming the nucleus by equations of form (46.4) and (46.7). Writing  $no = r^2$ , where  $o$  is the area occupied by an adsorbed atom,

$$r_c = \gamma o / (g^v - g^s). \quad (60.18)$$

In an exactly analogous manner to that used on p. 427, a nucleus of radius  $r$  may be shown to have a vapour pressure  $p$  given by

$$\ln(p_r/p_\infty) = \gamma o / r k T \quad (60.19)$$

and the excess free energy of a critical size nucleus is

$$\Delta G_c = \pi \gamma r_c = \pi \gamma^2 o / (k T \ln i). \quad (60.20)$$

Note that the excess free energy of a two-dimensional nucleus is one-half of the edge energy, just as that of a three-dimensional nucleus is one-third of the surface energy. The area  $o$  in the above equations is often replaced by the atomic volume  $v = o d_i$ , where  $d_i$  is the step height for a monolayer on the surface. Consider alternatively a multilayer nucleus with  $d = m d_i$ : the critical radius  $r_c$  is unchanged provided  $\gamma$  is proportional to  $d$ , i.e. provided the energy per unit area of a step is independent of the step height. In this case, eqn. (60.20) shows that the free energy of the critical size nucleus is  $m$  times larger than that of a monolayer nucleus. Even if the step energy is not proportional to  $d$ , it must increase with increasing  $d$ , so that  $G_c$  is always smallest for a monolayer nucleus, as is intuitively obvious.

If the nucleus is square, as we might expect for example in a cubic crystal, the above equations are modified slightly. For a nucleus of side  $l$ , the total change in free energy is

$$\Delta G = l^2 (g^s - g^v) / o + 4 l \gamma \quad (60.21)$$

and the critical nucleus is of size

$$l_c = 2 \gamma o / (g^v - g^s). \quad (60.22)$$

This is in equilibrium with a vapour pressure  $p$  such that

$$\ln i = 2 \gamma o / l k T \quad (60.23)$$

and the excess free energy of the critical size nucleus

$$\Delta G_c = 2 l_c \gamma = 4 \gamma^2 o / (k T \ln i) \quad (60.24)$$

is once again one-half of the edge free energy.

For the three-dimensional nucleation of a liquid from the vapour, isotropic surface energy was assumed so that the equilibrium shape is a sphere. It has already been shown that the equilibrium shape for a crystal is given by an extended form of Wulff's theorem. The above equations for a two-dimensional nucleus will now be generalized to obtain the equilibrium shape for any specified variation of edge energy with direction. In two dimensions, Wulff's theorem shows that the equilibrium shape will be given by the inner envelope of the polar free energy diagram, in which the distance of the origin from any edge is proportional to the edge free energy per unit length.

If the equilibrium shape contains an edge of length  $l(\theta)$  at a distance  $z(\theta)$  from the origin,  $\theta$  being the angle from any fixed direction, it follows that  $z(\theta) = z'\gamma(\theta)$ , where  $\gamma(\theta)$  is the free energy of the edge  $l(\theta)$  per unit length. The total area of the nucleus is

$$\sum (1/2) z(\theta) l(\theta) = \sum (1/2) z' \gamma(\theta) l(\theta).$$

As  $z'$  is varied, the size but not the shape of the nucleus changes, so that we may write  $l(\theta) = z' l_0(\theta)$ , where  $l_0(\theta)$  does not change with the size of the nucleus. The total edge free energy of the nucleus is  $\sum l(\theta) \gamma(\theta) = \sum z' l_0(\theta) \gamma(\theta)$ . The change in free energy resulting from the formation of this two-dimensional nucleus is thus

$$\Delta G = \sum (1/2) z'^2 \gamma(\theta) l_0(\theta) (g^s - g^v) / v_0 + \sum z' l_0(\theta) \gamma(\theta) \quad (60.25)$$

and  $\partial \Delta G / \partial z' = 0$  when

$$z' = v_0 / (kT \ln i), \quad (60.26)$$

and the equilibrium size and shape of the critical nucleus at a supersaturation  $i$  are therefore given by

$$z(\theta) = v_0 \gamma(\theta) / (kT \ln i). \quad (60.27)$$

In deriving eqn. (60.26), it has been assumed that the equilibrium nucleus is polygonal in shape. As already discussed on pp. 156-160, this may not be true, and at finite temperature the edges may be rounded. In the general case, Burton *et al.* (1950-1) showed that

$$z(\theta) / \gamma(\theta) = \text{const.} = v_0 / (kT \ln i) \quad (60.28)$$

where  $z(\theta)$  is now the pedal to the equilibrium shape (see p. 159). Thus eqn. (60.26) is obtained again, but with a more general significance;  $z(\theta)$  is no longer the radius vector from the origin, and as previously emphasized the equilibrium shape is not normal to the radius vector at all points. Actual calculations of the equilibrium shape for the {001} face of Kossel crystal (Burton *et al.*, 1949) show that this is a square with sides along the  $\langle 100 \rangle$  directions of the face at low temperatures, but the corners become progressively more rounded as the temperature is raised, until the nucleus is nearly circular close to the melting point.

At points where a circle of radius  $r$  with the origin as centre touches the boundary of the equilibrium nucleus, eqn. (60.27) corresponds to eqn. (60.18). The excess free energy of the critical nucleus is given by

$$\Delta G_c = \sum v_0 l_0(\theta) \gamma(\theta) \gamma / (2kT \ln i) \quad (60.29)$$

for a polygonal nucleus, and in the more general case

$$\Delta G_c = \int v_0 \gamma \gamma(\theta) / 2kT \ln i \cdot dl. \quad (60.30)$$

For a surface with fourfold symmetry,  $\Delta G_c$  will lie between the values for a circular and a square nucleus. In all cases we may write

$$\Delta G_c = 4C_{22} \gamma_2 v_0 / (kT \ln i), \quad (60.31)$$

where  $\gamma_2$  is the effective edge energy and  $C_{22}$  is close to unity. ( $C_{22}$  is between one and  $(1/4)\pi$  for a surface with fourfold symmetry.)

The rate of formation of two-dimensional nuclei per unit area of surface will be given by an expression of the form

$${}^S I = C_{23} \exp(-\Delta G'_c/kT) \quad (60.32)$$

where, by analogy with eqn. (48.13),  $C_{23}$  contains the number of atoms participating in the nucleation, the frequency with which a critical nucleus becomes supercritical, a Zeldovich factor and possibly thermal and statistical (Lothe-Pound) correction factors. The growth and shrinkage of nuclei may take place both by direct interchange with vapour atoms and by interchange with ad-atoms through surface diffusion, but the latter process is usually predominant (Burton and Cabrera, 1949; Pound *et al.*, 1954). If the Lothe-Pound correction is ignored, the areal "equilibrium" distribution function of embryos of  $n$  atoms is thus

$${}^S N_n = {}^S N_1 \exp(-\Delta G'_n/kT) \quad (60.33)$$

where  ${}^S N_1 = \alpha^S/o = i\alpha_0^S/o$  is the surface concentration of ad-atoms. For the case of ad-atoms and embryos localized on a substrate, Lothe and Pound (1962) and Hirth and Pound (1963) consider that the additional statistical factor [see eqns. (46.17) and (47.3)] will give, in place of eqn. (60.33),

$${}^S N_n = {}^S N_o \exp(-\Delta G'_n/kT) \quad (60.34)$$

where  ${}^S N_o = (1/o)$  is the surface density of lattice sites and  $\alpha^S$ , as in the analogous case of three-dimensional nucleation, has disappeared from the equation. The concentration of embryos given by eqn. (60.34) is larger than that given by eqn. (60.33) by a factor which depends mainly on  $w^k$  [eqn. (60.1)] and is thus typically of the order of  $10^3$ , but may vary from 10 to  $10^6$  in extreme cases.

The kinetic factor  $q_0 O_c$  of the three-dimensional theory must be replaced by the rate at which ad-atoms cross the periphery of the two-dimensional nucleus of critical size. Because critical nuclei are in equilibrium with the ad-atom population, in the first approximation the number of ad-atoms within one jump distance of a circular nucleus of radius  $r_c$  is  $2\pi r_c \alpha^S/o$ . The rate at which an adjacent ad-atom is added to the nucleus is approximately  $v \exp(-\varepsilon^S/kT)$  and the product of these two quantities gives the required kinetic factor. The Zeldovich factor allows for the decrease in the concentration of critical nuclei below that given by eqns. (60.33) or (60.34) and for shrinkage as well as growth of supercritical nuclei, and is calculated in a very similar way to that used in Chapter 10 [eqn. (48.9)] for the three-dimensional problem. This gives for a circular nucleus (Hirth, 1959)

$$\Gamma_{\zeta}^S = (\Delta G'_c/4\pi k T n_c^2)^{1/2} = kT(\ln i)^{3/2}/2\gamma o^{1/2} \quad (60.35)$$

using eqn. (60.20) and the relation  $n_c = \pi r_c^2/o$ . Thus, ignoring any possible thermal factor, the pre-exponential factor in the nucleation rate is

$$C_{23} = \{(\alpha^S)^2 r_1 (\ln i)^{1/2} v/o^{3/2}\} \exp(-\varepsilon^S/kT) \quad (60.36)$$

according to eqn. (60.33) and

$$C_{23} = \{(\alpha^S r_1 (\ln i)^{1/2} v / o^{3/2})\} \exp(-\varepsilon^S / kT) \quad (60.37)$$

if eqn. (60.34) is used. For nuclei of shapes varying from circular to square, the kinetic and Zeldovich factors are multiplied by  $(4C_{22}/\pi)$  and  $(\pi/C_{22})^{1/2}$  respectively, so that  $C_{23}$  is multiplied by  $(4C_{22}/\pi)^{1/2}$ .

Two different models may now be used to relate the growth rate of the crystal face to the nucleation rate  $^S I$ . If the step velocity is sufficiently large in relation to the nucleation rate, each successive layer will be formed by the spreading of the first successful nucleus, and there will then be a pause until the next layer is nucleated. The growth rate is thus given by

$$\Gamma = O^S S I d = O^S d C_{23} \exp(-\Delta G'_c / kT) \quad (60.38)$$

where  $O^S$  is the area of the face and  $d$  is the step height. Frank (1952) estimated the maximum value of  $C_{23}$  to be  $\sim 10^{27} \text{ mm}^{-1} \text{ s}^{-1}$ , so that a growth rate of  $10^{-3}$  atom layers per second ( $\sim 1 \mu\text{m}$  per month) on a  $1 \text{ mm}^2$  face would require a value  $\Delta G'_c / kT \cong 70$ . Using typical values in eqn. (60.30), this means that a supersaturation of 25–50% is required to sustain even this very small growth rate. Moreover, the dependence of the nucleation rate on the degree of supersaturation is so critical that a small change in the latter corresponds to a very large change in growth rate. The rate of crystal growth should thus be zero below some critical supersaturation, and above this value should not be limited by the nucleation rate, so that eqn. (60.38) no longer applies. Rather lower values of the critical supersaturation are obtained if eqn. (60.37) is used for the nucleation rate, but the general conclusion that the growth rate should be effectively zero at supersaturation of a few per cent remains, and is contrary to most of the experimental evidence.

The second growth model applies when new layers nucleate before existing layers are completed, i.e. as the critical supersaturation of the first model is approached. This leads to so-called "birth and spread", "polynuclear" or "multilayer" models, first considered (independently) by Nielsen (1964) and by Hillig (1966). Several layers are now growing simultaneously at the peripheries of separately nucleated regions, but such growth ceases locally with the impingement (and possibly the coalescence) of two such regions in the same layer. The resulting topology is very close to that envisaged for growth of two-dimensional "islands" on substrates. A steady-state growth rate may be derived analytically in the limiting case of copious nucleation, but predictions about the initial transient behaviour, and the eventual steady-state growth when there are few nuclei per layer, generally require computer modelling (see, e.g., Gilmer, 1980). An important factor in models of multilayer growth is whether or not a growing layer is allowed to "overhang" an incomplete lower layer. Such overhangs are probably physically unrealistic and models in which they are not permitted are described as "solid-on-solid" (SOS) models.

A nucleus originating at time  $\tau$  and growing with a velocity  $u_\infty$  will cover an area  $4C_{22}u^2(t-\tau)^2$  at a later time  $t$ . Between times  $\tau$  and  $\tau + \delta\tau$ , the number of new nuclei formed on the various partially completed layers will be  $O^S I \delta\tau$  as the total area exposed to

the vapour always remains  $O^S$ . After an initial transient which will depend on the starting configuration, a steady-state growth rate may be attained. Provided  ${}^S I/u_\infty$  is large enough to ensure that many two-dimensional nuclei contribute to the growth of every layer, the average time  $t$  to form each successive layer is now independent of  $O^S$  and is given according to Hillig (1966) by

$$\int_0^t 4C_{22}u_\infty^2(t - \tau)^2 {}^S I d\tau \cong 1. \tag{60.39}$$

The growth rate is thus

$$Y = d(4C_{22}u_\infty^2 {}^S I/3)^{1/3} = d(4C_{22}C_{23}u_\infty^2/3)^{1/3} \exp(-\Delta G'_c/3kT) \tag{60.40}$$

We note in passing that the conclusion that the effective activation energy for growth controlled by two-dimensional nucleation is  $\Delta G'_c/3$  was first reached by Burton and Cabrera (1949). Other derivations (Gilmer, 1980; Obretenov *et al.*, 1986) introduce an additional numerical factor into eqn. (60.40); as this is close to unity ( $\sim 0.97$  according to Obretenov *et al.*, 1989), we may simply incorporate it into  $C_{22}$ . Hillig was concerned specifically with growth from the liquid phase and he concluded that the pre-exponential term in the growth rate is proportional to the thermodynamic driving force (i.e. to the supercooling) to the power  $5/6$ . As  $u_\infty \propto i - 1 \cong \ln i$  [eqn. (60.5)] and  $C_{23} \propto (\ln i)^{1/2}$ , the pre-exponential factor in eqn. (60.40) similarly varies as  $(\ln i)^{5.6}$  (Bennema and Van Leuwen, 1975). (With the present model, the variation is strictly with  $i(\ln i)^{5.6}$  because of the factor  $\alpha^S$  in eqn. (60.34), but the additional variation with  $i$  is negligible.) De Haan *et al.* (1974) have shown by computer simulations using SOS models that eqn. (60.40) gives a good description of growth by two-dimensional nucleation provided  $2E_1/kT > 1.8$  and  $\ln i < 0.5$ . As mentioned on p. 154, surface roughening occurs for  $2E_1/kT < 1.6$ , and the growth is then linear in  $\ln i$ .

Clearly, there must be a growth regime for perfect crystals which is intermediate between mononuclear layer (MN) growth with one nucleus per layer and polynuclear layer (PN) growth with many two-dimensional nuclei per layer. As  $O^S {}^S I$  nuclei are formed in unit time, eqn. (60.39) may be generalized to

$$Y = (O^S {}^S I d)/N_L \tag{60.41}$$

where  $N_L$  is the average number of nuclei contributing to the formation of each layer. Thus the growth rate is expressed simply in terms of  $N_L$  and Obretenov *et al.* (1989) suggested a simple linear interpolation for the transition from one nucleus to many nuclei per layer by noting that  $x = ({}^S I/u_\infty)^{2/3} O^S$  is a dimensionless combination of the three principal parameters determining the growth rate and hence should be related to the number of nuclei per layer. By assuming  $N_L = 1 + a_1 x$ , and equating eqns. (60.40) and (60.41) for large values of  $x$ , it follows that  $a_1 = (4C_{22})^{1/3}$  and

$$Y = \frac{d {}^S I O^S}{1 + O^S (4C_{22})^{-1/3} ({}^S I/u_\infty)^{2/3}} \tag{60.42}$$

Obretenov *et al.* tested this equation by means of a Monte Carlo simulation in which the parameters were so chosen that the number of nuclei per layer varied between two and 80. The average number of nuclei involved in the growth of the sixth to the fifteenth monolayers (thus avoiding the initial transient) was found to be linear in  $x$  and the growth rate followed eqn. (60.42) quite closely. Previous simulations by Nielsen (1964) and by Viola *et al.* (1979) do not cover a wide range of  $x$ ; Nielsen's results do not agree well with eqn. (60.42), but those by Viola *et al.* appear to be in good agreement with this prediction. In contrast to earlier work, it is also emphasized that the MN assumption requires special values of the parameters and that, under most conditions, PN growth is to be anticipated. Also, if there are more than three nuclei per layer ( $x \geq \sim 5$ ), eqn. (60.40) is a better approximation than eqn. (60.39), so that if eqn. (60.42) cannot be used, eqn. (60.40) will provide the best approximation in almost all circumstances. There is, however, a range of high driving forces for which the above model does not apply and for which atomistic calculations are essential; this is considered further in Section 62.

In the majority of the cases which have been studied experimentally, there is a large discrepancy between the predictions of the theory and the observation that appreciable growth may take place at supersaturations of the order of 1%. Once the stepped faces have disappeared, at least one fresh two-dimensional nucleus is required for every atomic plane of a perfect crystal, and effects of this kind cannot therefore be attributed to heterogeneous nucleation on impurities, which is usually the reason for discrepancies between theory and experiment for three-dimensional nucleation. The solution of this problem lies in the recognition that the theory as developed above applies only to the growth of a perfect crystal. An imperfect crystal grows in a different manner which was predicted theoretically by Frank and has been amply confirmed by experiment.

## 61. REAL CRYSTALS: FRANK'S THEORY OF CRYSTAL GROWTH

The theory of the growth of real crystals, due to Frank, assumes that dislocations are always present, and that these remove the necessity for two-dimensional nucleation on close-packed surfaces. As previously noted, when a simple screw dislocation emerges at a crystal face, a step is produced across part of the face. This step (for example the line  $PQ$  in Fig. 7.4) has a height equal to the Burgers vector if the screw dislocation is normal to the crystal face. More generally, any dislocation line which has a component of its Burgers vector perpendicular to the surface at which it emerges will produce a step on that surface. The point of emergence will form one end of the step, and the height of the step will be equal to the component of the Burgers vector normal to the emergent face. The importance of this result is that, for a dislocated crystal, an atomic step may end within the crystal face, in contrast to the statement on p. 148.

Figure 13.8 shows the end of a simple screw dislocation in the building-block model (Kossel crystal). This step is self-perpetuating; no matter how many layers of atoms are deposited on the crystal face, the step will persist. For a single screw dislocation emerging

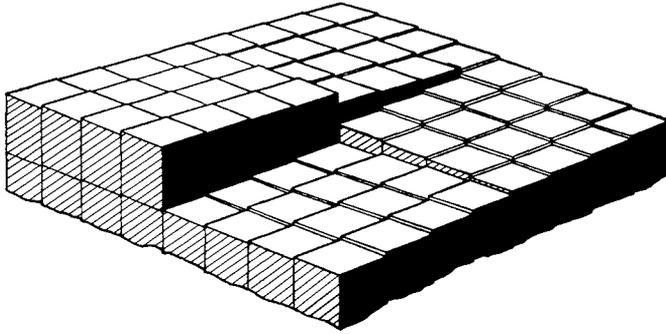


FIG. 13.8. Model of emergent screw dislocation (after Read, 1953).

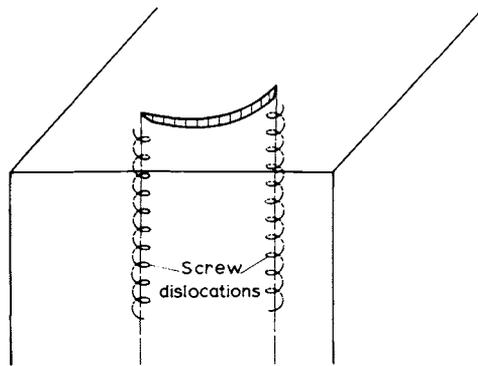


FIG. 13.9. Surface step joining two opposite screw dislocations.

at the centre of a face, the crystal will be able to grow up a “spiral staircase” by continually depositing atoms at the exposed step, which rotates continuously about the point where the dislocation emerges. The structure of the crystal containing this dislocation is one continuously connected atomic plane, wrapped around itself in the form of a helicoid. The structure of a crystal surface normal to two emergent screw dislocations of opposite sign is shown in Fig. 13.9. There is a step of height equal to the Burgers vector running from the end of one dislocation to the end of the other. In a face in which a large number of dislocation lines emerge, they will mostly be linked together in pairs in this way. Any isolated lines, or excess lines of one sign, will form steps running to the edge of the surface.

The kinetics of the growth of a dislocated crystal were developed by Burton *et al.* (1950–1) in considerable detail, and have been further refined by subsequent workers. If a dislocated face is in equilibrium with the vapour, the step from a single dislocation will run in a straight line from the dislocation axis to the edge of the face. When the vapour is supersaturated, atoms are deposited along this step which thus advances along the crystal face with finite velocity. If the step remains straight, its velocity will be given by eqn. (60.5).

Because, however, the end of the step is fixed on the dislocation line, the whole step cannot advance in this way or a sharp corner would be produced. In fact a step with radius of curvature  $\rho_c = r_c$  [eqn. (60.18)] will have zero velocity of advance, as this step is in (unstable) equilibrium with the vapour. The velocity of the step will therefore become smaller and smaller as the dislocation end is approached, and the step will take up a curved form, with radius of curvature continuously decreasing, but remaining everywhere greater than  $\rho_c$ . We introduce the new notation  $\rho_c$  in order to emphasize that, in the spiral step which we are about to consider, this is a critical value of the radius of curvature rather than of the coordinate  $r$  which is used to define the shape of the spiral (see Fig. 13.12). Also, if  $\rho_c$  is defined in this way, cases of undersaturation as well as supersaturation may be included by allowing  $\rho_c$  to have negative as well as positive values.

A curved step advances more slowly than a straight step because the evaporation rate is greater by a factor  $i'$ , where  $i'$  is the supersaturation ratio giving the pressure of vapour which would be in equilibrium with the particular step considered. For a curved step, we thus have instead of eqn. (60.5)

$$u = 2\beta(i - i')x^S \exp(-\Delta h^{sv}/kT). \quad (61.1)$$

For small supersaturations, eqn. (60.19) gives

$$\ln i \cong i - 1 \cong \gamma o / \rho_c kT \quad (61.2)$$

Similarly

$$i' - 1 \cong \gamma o / \rho kT \quad (61.3)$$

where  $\rho$  is the radius of curvature of the step. We thus obtain

$$i - i' = (i - 1)[1 - (\rho_c/\rho)] \quad (61.4)$$

and

$$u = u_\infty [1 - (\rho_c/\rho)] \quad (61.5)$$

as we should expect, as  $u = 0$  for  $\rho = \rho_c$  and  $u = u_\infty$  for  $\rho = \infty$ .

Figure 13.10 shows the shape of an initially straight step growing in supersaturated vapour at successive periods of time, the shading indicating the higher side of the step. It can be seen that the step will rapidly wind up into a spiral shape, centred on the end of the dislocation, and with a curvature everywhere less than  $\rho_c^{-1}$ . When the curvature at the centre reaches this critical value, the spiral will assume a stationary form, and as atoms are added at the growth step after this, the whole spiral will rotate with uniform angular velocity. The growing face of the crystal will thus have the form of a spirally terraced hill, the height of the terraces in single lattice structures being one interatomic distance for pure screw dislocations with Burgers vectors equal to the shortest lattice vector, and more for dislocations of multiple strength. There is clearly a close analogy to the operation of a single-ended Frank-Read source (p. 248) and an even closer analogy to a Bardeen-Herring source (p. 250).

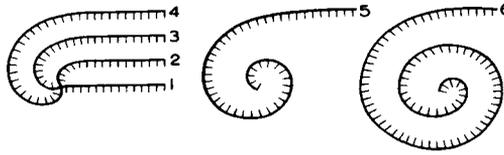


FIG. 13.10. Development of spiral step structure.

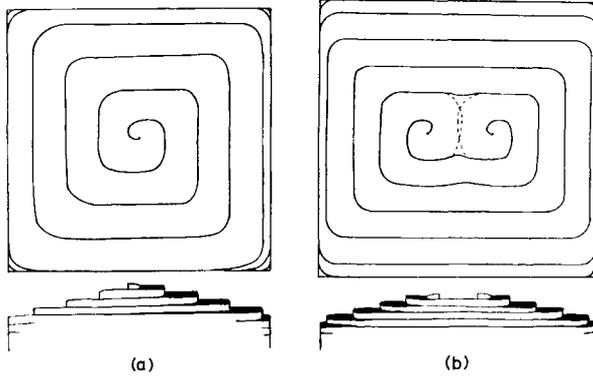


FIG. 13.11. Growth pyramid due to a single dislocation (a) and a pair of dislocations (b) (after Burton *et al.*, 1950-1).

Steps on a crystal face due to right- and left-hand dislocations will grow in a similar manner, provided the dislocation ends are separated by a distance greater than  $2\rho_c$ ; it is evident that no growth will occur if this is not the case. The terrace step spreads out on one side, thus forming a closed terrace loop which grows outwards and a short connecting step which repeats the process. This process is analogous to the action of a double-ended Frank-Read or Bardeen-Herring source. From such a pair of opposite dislocations, we thus obtain closed terraces on a flat cone, instead of the continuous terrace of the isolated dislocation. The structures of crystal faces growing by isolated and paired dislocations are shown diagrammatically in Fig. 13.11.

The rate of growth of a crystal face containing one screw dislocation will be given by  $\omega d/2\pi$ , where  $\omega$  is the angular velocity of the spiral in the steady state and  $d$  is the step height. In order to calculate this velocity, the form of the spiral satisfying eqn. (61.5) must be obtained. The equation of a spiral is most conveniently written in polar coordinates  $r$  and  $\theta$ ; each value of  $r$  uniquely defines  $\theta$ , but for any value of  $\theta$  there is an infinity of values for  $r$ . The whole spiral, and its coordinate system, is considered to be rotating with angular velocity  $\omega$ . At a point  $r$  of the curve, Fig. 13.12, the angle  $\phi$  between the radius vector and the tangent to the curve is given by  $\sec \phi = [1 + r^2(d\theta/dr)^2]^{1/2}$ . If the angular velocity is  $\omega$ , the linear velocity perpendicular to the step at the point  $r$  is

$$u(r) = \omega r \cos \phi = \omega r [1 + r^2(d\theta/dr)^2]^{-1/2} \quad (61.6)$$

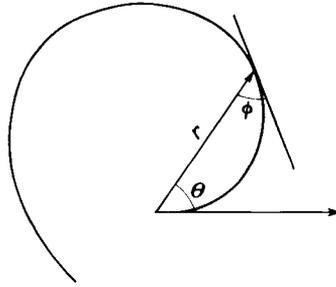


FIG. 13.12. The equation of a growth spiral.

Finally, a relation is needed between the radius of curvature and the polar coordinates  $r$  and  $\theta$  in order to link eqns. (61.5) and (61.6). This relation is

$$\rho = \frac{\{1 + r^2(d\theta/dr)^2\}^{3/2}}{2(d\theta/dr) + r^2(d\theta/dr)^3 + r(d^2\theta/dr^2)} \quad (61.7)$$

The correct form of the spiral is found by solving eqns. (61.5), (61.6) and (61.7). The simplest type of spiral is the Archimedian spiral, for which  $r/\theta$  is constant, and a spiral having radius of curvature  $\rho_c$  at the centre will have the equation

$$r = 2\rho_c\theta \quad (61.8)$$

From eqn. (61.6), taking the limit for large  $r$ ,

$$\omega = u/2\rho_c \quad (61.9)$$

Equations (61.8) and (61.9) do not satisfy eqn. (61.7) and cannot be the exact solution. Moreover, as first emphasized by Cabrera and Levine (1956), the stress field of the dislocation may have a considerable effect on the step velocity and hence on the shape of the spiral. This effect arises because the strain energy of the dislocated solid increases as its volume increases. The additional strain energy may be regarded as a further contribution to the driving force, opposing growth and aiding evaporation or dissolution. If  $w(r)$  is the strain energy per unit volume added by growth at a step distance  $r$  from the dislocation centre, the equivalent change  $\omega dw(r)$  in the driving force per atom may be written from eqn. (60.18) as  $w(r)[\rho_c d/\gamma](g^v - g^s)$ , so that eqn. (61.5) for the step velocity has to be modified to

$$u = u_\infty [1 - (\rho_c/\rho) - (\rho_c d/\gamma)w(r)] \quad (61.10)$$

Combining eqns. (61.6), (61.7) and (61.10), the differential equation for the spiral may now be written in the form

$$\frac{d\phi}{ds} = 1 - \frac{\omega(s)d\rho_c}{\gamma} \sec \phi - \frac{\tan \phi}{s} - \omega_1 s \quad (61.11)$$

where, following Cabrera and Levine (1956) and van der Hoek *et al.* (1982b),

$$s = r/\rho_c \quad (61.12)$$

and

$$\omega_1 = \omega\rho_c/u_\infty \quad (61.13)$$

have been introduced.

The strain energy density  $w(r)$  defined above will differ from the energy density  $(2r)^{-1}(\partial'W_s/\partial r)$  around a dislocation in the interior of a solid but, as discussed by van der Hoek *et al.* (1982a), the “interior” or “bulk” energy is a reasonable approximation to the actual energy. This gives in the approximation of isotropic elasticity [see eqn. (30.19)]

$$w(r) = 1/2 Bb/4\pi r^2 = r_0\gamma/r^2d \quad (61.14)$$

where the “Frank radius”  $r_0$  is defined by this relation. Equation (61.14) is only valid outside the core region of the dislocation; inside the core, van der Hoek *et al.* (1982a) assume a constant energy density  $w(0)$ , and they express the total energy density for all  $r$  as an analytical function

$$w(r) = w(0)/[1 + (r^2/r_h^2)] = (r_0\gamma/r_h^2d)/[1 + (r^2/r_h^2)] \quad (61.15)$$

where the core region is effectively defined by the “Hookean” radius  $r_h$ , a rough estimate of which may be obtained by equating  $w(0)$  with the latent heat of melting.

One effect of the stress field, first pointed out by Frank (1951), is that dislocations of large Burgers vector emerging at a free surface may have hollow cores, and this in turn influences the shape of the growth or dissolution spirals. Using the above analytical stress field, van der Hoek *et al.* (1982a) found the thermodynamic conditions for stable hollow cores. For large stress fields (i.e. values of  $r_0/r_h > 8\sqrt{3}/9$ ), steep etch pits are stable at large undersaturations, but when the undersaturation is less than a critical amount (given approximately by  $r_0/\rho_c = -1/4$ ), the stable configuration is a hollow core. The range of stability of the hollow core extends to a second critical driving force which may be a smaller undersaturation or a supersaturation which increases with increasing stress field, i.e. increasing  $r_0/r_h$ . The hollow core is thermodynamically stable under growth conditions (supersaturation) provided  $r_0/r_h$  exceeds about two, but above the second critical driving force the core will be closed. The range of undersaturation–supersaturation over which the hollow core is thermodynamically stable increases as the stress field (and hence the Burgers vector) increases and this effect is due to the dependence of the second critical supersaturation on  $r_0/r_h$ ; the undersaturation corresponding to the first critical condition is nearly constant. Finally, the “phase diagram” for the core configuration thus predicted depends on the assumption of equilibrium and may not be achieved under certain kinetic conditions; for example, if the growth or dissolution rate inside the predicted hollow core is respectively greater or smaller than the rate outside it.

The radius of the hollow core  $r_{hc}$  is always greater than  $r_h$  and, because in the approximation of eqn. (61.15)  $r_h$  is of the order of magnitude of the Burgers vector, it follows that visible hollow cores will be expected only from dislocations with large Burgers

vectors, as observed experimentally. If  $r_h = 0$ , which corresponds to the assumption of linear elastic behaviour outside an empty core, then  $r_{hc} \rightarrow r_0$  as  $\rho_c \rightarrow \infty$ , i.e. no supersaturation or undersaturation, and this corresponds to the original result of Frank (1951). At finite supersaturation or undersaturation, the assumption  $r_h = 0$  leads to

$$r_{hc} = \frac{1}{2} \rho_c [-1 + (1 + 4r_0/\rho_c)^{1/2}] \quad (61.16)$$

This result was first given by Cabrera and Levine (1956).

The solution of eqn. (61.11) requires the specification of the boundary conditions at  $s=0$  and  $s=\pm\infty$  for normal dislocations and also at  $s=c=r_{hc}/\rho_c$  for hollow cores. The first two conditions are readily seen to be  $\phi(0)=0$  and  $\phi(\pm\infty)=\pm\frac{1}{2}\pi$ , where the positive sign refers to growth spirals and the negative sign to dissolution. Growth centred on a hollow core dislocation will produce a hillock around the hollow core, so that the surface slope changes sign at  $s=c$ , and this implies a change in sign of  $\phi$  so that  $\phi(c)=-\frac{1}{2}\pi$ . For dissolution, in contrast, there will be no change in sign of  $\phi$  as to produce the hollow core simply requires an increase in the surface slope near  $s=c$ . Hence the condition  $\phi(s)=-\frac{1}{2}\pi$  is valid for both negative and positive values of  $s$ . In the absence of stress, the solutions are symmetrical for growth and dissolution and in all cases the turns of the spiral become equi-spaced at large distances from the centre. The separation of the steps at large (strictly infinite) distances from the dislocation is given by

$$\Delta r_\infty = 2\pi\rho_c/\omega_1 \quad (61.17)$$

Equation (61.9) corresponds to  $\omega_1 = 0.5$  and  $\Delta r_\infty = 4\pi\rho_c$ . Burton *et al.* showed that a better analytical approximation to the solution of eqn. (61.15) is

$$s + \ln\left[1 + s/\sqrt{3}\right] = 2(1 + \sqrt{3}\theta) \quad (61.18)$$

and the corresponding value of  $\omega_1$  is  $\sqrt{3}/(1 + \sqrt{3}) = 0.315$ , giving  $\Delta r_\infty = 19.8\rho_c$ . A numerical solution to the differential equation in the absence of stress was first obtained by Ohara and Reid (1973) and their result,  $\omega_1 = 0.331$ ,  $\Delta r_\infty = 18.98\rho_c$ , was confirmed by Swendsen *et al.* (1976) and by Gilmer (1976) using Monte Carlo techniques.

Analytical approximations to the solution of eqn. (61.11) with the stress field included have been given by van der Hoek *et al.* (1982b) for the limiting cases  $s \rightarrow 0$ ,  $s \rightarrow c$  and  $|s| \rightarrow \pm\infty$ , and these authors derived the shapes of the whole spirals by phase plane analysis. The results indicate that three kinds of spiral are possible, namely (i) global spirals running from the dislocation centre ( $r=0$ ) to infinity, (ii) inner spirals from  $r=0$  to  $r=r_{hc}$  and (iii) outer spirals from  $r=r_{hc}$  to infinity. If the hollow core is both thermodynamically and kinetically possible, only types (ii) and (iii) occur. A striking prediction is that a growth spiral emanating from a dislocation with a large Burgers vector will show a change in curvature from negative to positive as it moves outwards from the centre, and this may apply both to cases where the hollow core is stable or where (at higher supersaturation) it is no longer stable.

Important results from this analysis are that the dissolution or growth rates remain finite over the whole range of under- and supersaturations, that the stress field opposes

growth slightly but aids dissolution considerably, and that eqn. (61.17) continues to be valid in the presence of a stress field, but with a value of  $\omega_1$  which may be very different from the stress-free value. At high undersaturations, the authors predict a closely spaced dissolution spiral, whilst at low undersaturations there are hollow core inner and outer dissolution spirals. Similarly, at low supersaturations there are hollow core inner and outer growth spirals, the outer spiral having a change of curvature after it leaves the hollow core, whilst at higher supersaturations the hollow core closes, leaving first a growth spiral which still has a region of reversed curvature near the centre and finally a widely spaced Archimedean type spiral. The step spacing  $\Delta r_\infty$  at large distances from the dislocation may vary widely from its zero stress value of  $19\rho_c$  up to about  $100\rho_c$  for growth and from  $-19\rho_c$  down to about  $-4\rho_c$  for dissolution.

When a steady-state spiral has been obtained, a crystal surface growing from an isolated dislocation will have a growth rate given by

$$Y = \omega d/2\pi = \omega_1 u_\infty d/2\pi\rho_c = u_\infty d/\Delta r_\infty \quad (61.19)$$

As  $\rho_c$  is proportional to  $j^{-1}$  [eqn. (61.2)], it follows that the growth rate is proportional to  $\omega_1 j^2$  provided that  $u_\infty$  is given by eqn. (60.5). If the stress field of the dislocation is very small, the predicted growth rate is proportional to the square of the supersaturation, as in the original Burton–Cabrera–Frank paper, as  $\omega_1$  then has the constant value of 0.33. The calculations show that the growth rate for any given supersaturation decreases with increasing stress field expressed as a variation in the value of  $r_0/r_h$  from zero (zero field,  $r_h = \infty$ ) to four; the rates in the latter case and at all higher stress fields are indistinguishable from those for the maximum stress field,  $r_0/r_h = \infty$ . The behaviour on evaporation (or dissolution) is different. There is a very large change in the evaporation rate on passing through the critical undersaturation of  $r_0/\rho_c = -1/4$  where, as noted above, the equilibrium configuration for stress fields  $r_0/r_h \geq 1.56$  changes from a hollow core to a steep etch pit. However, for  $r_0/r_h = 0$  or 1 (where a hollow core is not possible), the evaporation rate varies smoothly with undersaturation at all stresses.

The variation of growth rate with supersaturation will clearly be different under conditions where the successive turns of the spiral step compete with each other for the atomic flux, i.e. when the distance  $\Delta r_\infty$  is less than  $2x^s$ . If eqn. (60.6) is valid for a curved step with the distance between steps  $y$  replaced by  $\Delta r_\infty$ , the expression for the growth rate becomes

$$Y = d\beta v \exp(-\Delta h^{sv}/kT) [(\omega_1/\pi)(x^s kT/\gamma\sigma)] j^2 \tanh[(\pi/\omega_1)/(\gamma\sigma/x^s kTj)] \quad (61.20)$$

where we have substituted for  $\rho_c$  from eqn. (61.2) and written  $\ln i = i - 1 = j$  as we are dealing with relatively small supersaturations. This expression can be written more concisely if the symbol  $j_1$  is used for  $\pi\gamma\sigma/\omega_1 x^s kT$ . The growth rate is then

$$Y = d\beta v \exp(-\Delta h^{sv}/kT) (j^2/j_1) \tanh(j_1/j) = Y_m (j^2/j_1) \tanh(j_1/j) \quad (61.21)$$

When  $j \ll j_1$ ,  $\tanh(j_1/j) \rightarrow 1$  and when  $j \gg j_1$ ,  $\tanh(j_1/j) \rightarrow j_1/j$ . Equation (61.21) thus predicts, at least in the case where the stress field of the dislocation has little effect and  $\omega_1$  is thus effectively constant, that there will be a transition from a parabolic growth law in

which  $Y$  varies as  $j^2$  to a linear law as the degree of supersaturation is increased. There is a characteristic supersaturation  $j_1$  for this change, but the actual transition in kinetic behaviour is very gradual. When the stress field cannot be neglected, these simple functional dependencies of  $Y$  on  $j$  no longer apply because  $\omega_1$  (and hence  $j_1$ ) also varies with  $j$ .

The above results do not take account a so-called "back stress" effect first pointed out by Cabrera and Coleman (1963) which is significant at high degrees of under- and supersaturation. The supersaturation at the centre of a growth spiral will be reduced below the value  $j$  because of the flux of ad-atoms to the first turn of the spiral. This means that the curvature of the first turn of the spiral must be larger than  $\rho_c$ , and the spacing of the outer arms will be correspondingly altered. Equation (61.21) remains valid for  $j < j_1$ , but for  $j > j_1$  it is modified to

$$Y = Y_m \beta (j/j_1)^{1/3} \tanh(j_1/j)^{1/3} \quad (61.22)$$

so that the transition to the linear law  $Y = \beta Y_m$  occurs even more gradually. Surek *et al.* (1977) calculate that, under conditions of evaporation at large undersaturations, the limiting radius of curvature may be much larger than  $\rho_c$ .

The above equations for the shape of a growth spiral, and the consequent expressions for the growth rate, are obtained on the assumption that the velocity of a step depends only on its radius of curvature  $\rho$ . It is possible that either  $\beta$  or  $x^S$  is dependent on the orientation of the step, in which case the spiral will no longer have the "circular symmetry" of eqn. (61.10). For a cubic substance,  $x^S$  should be effectively independent of direction, and (see above) for growth from the vapour  $\beta$  will be very nearly unity, except for complex molecules. In other words,  $x^S \gg x_n$ , and no variation of velocity of steps should result from the fact that there are fewer kinks in close-packed steps. Metallic crystals growing from the vapour on steps formed from single dislocations or two opposite dislocations should therefore have circular growth spirals or closed circular growth loops respectively, as shown in Fig. 13.10. The growth rate is then given by eqn. (61.21) or (61.22).

For growth from liquid solution, or from the vapour phase in the presence of a carrier gas, it may no longer be valid to consider that surface diffusion is the dominant process, and volume diffusion through a liquid or gaseous layer around the surface may have to be included in the theory. In some cases, direct deposition of atoms on steps may then be a more rapid process than the diffusion of ad-atoms to the steps. Experimental results on the electrodeposition of silver indicate, for example, that direct deposition is at least 100 times faster than the surface diffusion process for all step separations (down to  $\sim 45$  nm) examined. More usually, surface diffusion remains dominant, but the resistance effect of the volume diffusion has to be added to that of the surface adsorption reaction, so that the volume and surface diffusion reactions are coupled (Gilmer *et al.*, 1971). One possible consequence at low temperatures is that the distance between kinks in a close-packed step may exceed the effective diffusion length and the factor  $\beta$  may become smaller than unity and vary with step orientation. The resultant dependence of step velocity on orientation leads to growth spirals of polygon shape, the polygons having at least the rotational

symmetry of the crystal face considered. The sides of the polygons are straight at low temperatures and become increasingly rounded as the temperature is raised. The corners appear to be macroscopically sharp, but actually must have radii of curvature in excess of  $\rho_c$ . There is much experimental evidence for polygon-shaped growth spirals (see, e.g., Frank, 1952; Verma, 1953).

Returning to the growth of metallic crystals from the vapour phase, notice that the growth rate has been calculated only for single dislocations and for two opposite dislocations emergent at a face. This is an unreal situation because if one dislocation is present, there may be many more. As discussed on p. 313, the dislocation density may vary from  $\ll 1$  to  $10^6 \text{ mm}^{-2}$ . A surprising result, however, is that an increase in dislocation density has a rather small effect on the growth rate, at least until a critical density is reached.

Consider first two pairs of dislocations of opposite signs, each of which will send out closed loops of steps. When two such loops meet on the same plane, they unite and the steps disappear. The number of steps passing any point in unit time will be the same as if only one pair existed, and the area of the face may be divided into two parts, each receiving steps from one of the pairs. Two such pairs of dislocations thus give a growth rate equal to that from one pair alone. The surface of the crystal will consist of a single growth cone, except near the dislocations where it will split into two cones. Unless the separation between the pairs is resolvable by the examining techniques used, this case will not be distinguishable from the appearance of a surface growing from a single dislocation pair.

Two single dislocations of opposite sign separated by a large distance produce an almost identical result. Once again, the area may be divided into two parts, fed with steps from the two respective spirals. There is a tendency for the rotations of the spirals to become synchronized. The growth rate and topology of the surface are very similar to those resulting from two pairs of dislocations.

Two single dislocations of the same sign produce slightly more complicated effects, depending on whether their separation is greater or less than  $2\rho_c$ . If they are far apart, the situation remains similar to that described above, but the locus of intersection is not a straight line but rather an S-shaped curve; this is illustrated in Fig. 13.13a. If the two dislocations are close together, however, the two pairs of spirals will not intersect, and the growth steps from each will reach the whole of the area. This is illustrated in Fig. 13.13b. The pair of cooperating dislocations send out steps twice as fast as a single dislocation, and the growth rate for small supersaturations (parabolic region) is doubled. This conclusion is strictly valid only for separations very much smaller than  $\rho_c$ ; for small but not negligible values of the separation, the growth rate is between one and two times that given by a single dislocation. In the same way, a group of  $n$  like dislocations close together can give a growth rate up to  $n$  times that of a single dislocation.

The growth rate of a crystal containing a group of dislocations can now be predicted. If the group contains equal numbers of dislocations of opposite signs, the growth rate will be given by eqns. (61.21) or (61.22), in which the gradual change to a linear growth law is still specified by the supersaturation  $j_1$ . However, because a pair of dislocations will not grow

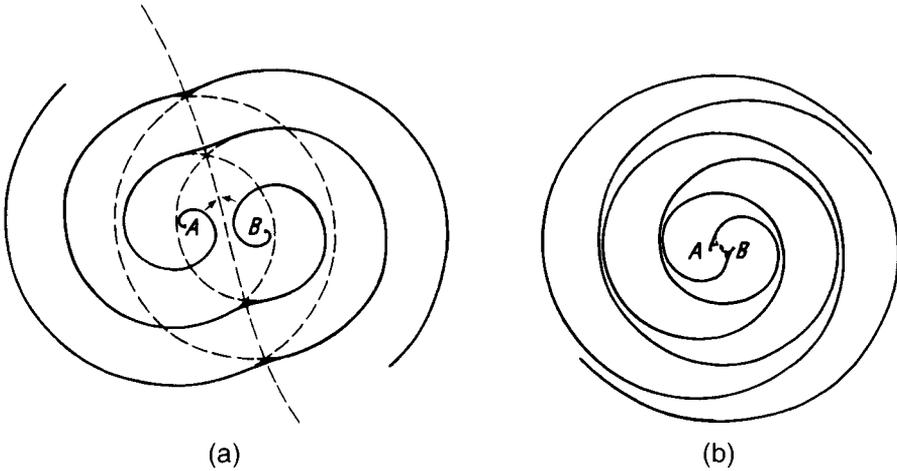


FIG. 13.13. Growth steps produced by a pair of like dislocations separated by (a) a distance greater than  $2\rho_c$  and (b) a distance less than  $2\rho_c$  (after Burton *et al.*, 1950 1).

unless they are separated by a distance greater than  $2\rho_c$ , there will be a second critical supersaturation  $j_2$  where the maximum distance between two dislocations joined by a step equals  $2\rho_c$ . There will be no growth centred on this group at supersaturations less than  $j_2$ .

A group of dislocations which contains an excess number of one sign may give, in the initial stages, a growth rate higher than those of eqns. (61.21) or (61.22) by a factor  $C_{24}$  which is between one and the number of unbalanced dislocations. Thus instead of eqn. (61.21) we have

$$Y = C_{24} Y_m (j^2/j_i) \tanh(j_i/C_{24}j) \quad (61.23)$$

from which it follows that the growth rate is higher at low supersaturations but tends at high supersaturations to the same linear law as is given by a balanced group or a single dislocation; the main difference is that the linear law is reached at lower supersaturations. As  $j_i$  is very small, observations on the parabolic region are, in any event, very unusual. The physical reason for the identical growth rate at higher supersaturations is, of course, that the spirals begin to compete for atoms so that each grows more slowly than it would in the absence of the others.

From this discussion, it follows that observations of the rate of growth of a crystal face cannot be used to determine the density of dislocations; essentially, they will only show whether or not there is a single growth-producing dislocation emergent at that face. The number of growth cones on a face gives a lower limit to the number of dislocations, but no reliable estimate, as separate cones from neighbouring dislocations will often be invisible. In addition, a growth cone with slightly higher supersaturation at its centre will tend to dominate the others, continually growing into their territory until there is essentially only one growth hill to be seen.

Soon after the original suggestion that crystal growth is catalysed by dislocations was made, experimental evidence of growth cones and pyramids (for growth from solution) was obtained for many substances. It was found that spiral steps and other predicted step configurations were often visible in the optical microscope, and the step height, measured by techniques such as multiple beam interferometry, was found in some cases to be the interplanar spacing of lattice planes parallel to the surface and in others to have much larger values. The reason why the small steps were visible was not always clear in the early experiments; there was usually an etching or "decorating" effect which in some cases was traced to vapour emanating from the plasticine used as specimen mounts.

The growth cones were originally found on the surfaces of organic and inorganic crystals, and many beautiful examples are shown in the early review article by Frank (1952) and in the book by Verma (1953). Amelinckx *et al.* (1957) made the first metallic observations on colloidal gold grown from solution, and Forty (1952) studied growth steps on magnesium and cadmium crystals deposited from the vapour phase. The step heights which he measured were only 0.26–0.52 nm, i.e. one or two interatomic distances. Forty and Frank (1953) made a detailed study of growth spirals on silver crystals, during the course of which they obtained the first experimental evidence for the movement of dislocations. Growth spirals have now been observed on many metallic crystals grown from the vapour or deposited electrolytically, and on many other crystals grown from solution.

The growth steps are often too finely spaced for detailed study to be possible by optical microscopy, and the greater lateral resolution provided by electron microscopy is then required. Studies of surface steps were greatly facilitated by the introduction of the gold decoration technique (Bassett, 1958) which has been much used by Bethge (1962, 1967), Keller (1969) and others in work on, for example, the growth and evaporation of NaCl crystals cleaved on {100} surfaces. The patterns they observed consisted mainly of circular spirals or closed loops of height  $\frac{1}{4}a$  (where  $a$  is the lattice parameter), together with a few polygon (nearly square) spirals or loops of height  $a$ . These two kinds of step originate from dislocations with Burgers vectors of types  $\frac{1}{2}a \langle 110 \rangle$  and  $a \langle 100 \rangle$  respectively. In addition to static studies of step structures, double decoration methods have been used to reveal the positions of a growth step both initially and after a measured time interval and thus to derive step velocities (Bethge *et al.*, 1968). The results obtained agree reasonably well with the theory of step growth inasmuch as the velocity over a certain range depended linearly on the supersaturation and exponentially on the temperature. Moreover, a plot of step velocity versus step separation at constant temperature and supersaturation, whilst not corresponding exactly to the tanh function of eqn. (60.6), showed the essential features of velocity first increasing with separation and then attaining a maximum value independent of separation [eqn. (60.5)] when the separation exceeds  $2x^S$ .

In the course of this work on the evaporation of cleaved NaCl surfaces in a high vacuum, Bethge and Keller observed that in some regions in which the steps formed closed loops there did not appear to be a visible central step, running from one dislocation to the other, which is characteristic of a double-ended source. The spacing of the steps in these regions was always considerably larger than that of conventional spiral or concentric

sources, and the authors concluded that the patterns originated from repeated two-dimensional nucleation at a favoured site at the centre of the pattern, rather than from the Frank mechanism. They further suggested that a suitable heterogeneous site might be the point of emergence of a dislocation which does not produce a step (i.e. an edge dislocation if it is normal to the emergent face). An alternative possibility is that the dislocations of a Frank double-ended source have subsequently migrated away, and although Keller showed that this process could be revealed by the decoration technique, the edge dislocation interpretation remained controversial until the work of Bauser and Strunk (1981). By an ingenious combination of surface step observations and high voltage transmission electron microscopy, these workers proved that the patterns without a central step are indeed centred on a dislocation with a Burgers vector parallel to the surface.

Dislocations which eliminate the need for two-dimensional nucleation are commonly called "screw", and dislocations which do not produce an indestructible step are often labelled "edge" but, as already noted, these are only special cases of the general condition relating the surface normal and the direction of the Burgers vector. Bauser and Strunk used the terms "transverse" and "longitudinal" for the two cases of Burgers vector parallel and non-parallel to the surface, whereas Frank (1951) suggests the use of the heraldic terms "couchant" and "rampant". It is clear from other experiments that dislocations with Burgers vector in (or almost in) the surface are not always sites for two-dimensional nuclei and, like the homogeneous nucleation considered in Section 60, we expect such nucleation to be very sensitive to the supersaturation. A dramatic example of the differential effects of different dislocations is shown in the X-ray topograph of a single crystal of an inorganic crystal grown from solution (Fig. 13.14). This technique gives information about the total dislocation content of a crystal which may thus be related to the growth form. The top (010) face is intersected by a single screw dislocation and has grown much more rapidly than the bottom (010) face which has no emergent dislocation. Dense bundles of dislocations emanating from the original seed crystal intersect the side (101) faces, but have not provided rapid growth of these faces because they all have Burgers vectors in the [010] direction.

Frank's prediction that dislocations with large Burgers vectors emerging at a free surface should have hollow cores was made quite early in the development of the theory (Frank, 1951) and, since that date, hollow dislocations (or at least unusually deep etch pits) have frequently been observed at the centres of growth spirals with large step heights in crystals such as SiC and CdI<sub>2</sub>. Sunagawa and Bennema (1981), however, were the first authors to attempt a detailed comparison of the predictions of the theory developed above with experimental measurements on growth spirals. Perhaps the most striking result obtained in this work is the visual confirmation (see Fig. 13.15) of the change in curvature from negative to positive in the central region of a spiral growth step originating from a dislocation in SiC which had a large Burgers vector and a hollow core. Similar but smaller reversals of curvature were observed on growth steps centred on dislocations (or dislocation groups) without hollow cores, but no such changes were observed on evaporation steps. Both these observations are consistent with the theory developed by van der Hoek *et al.* More detailed quantitative comparison of theory with experiment was

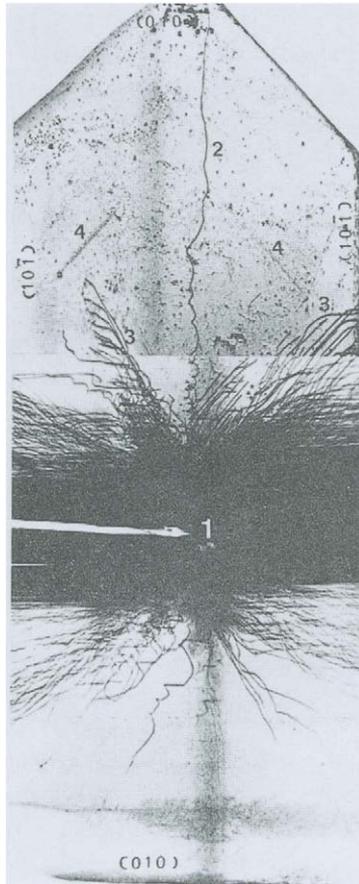


FIG. 13.14. X-ray topograph of solution-grown crystal of triglycine sulphate (after Israel et al., 1972): 1 = seed; 2 = screw dislocation; 3 = dislocations with  $b$  in IOT.

also attempted, for example by comparing the values of the Frank radius deduced (via  $\rho_c$ ) from observed values of  $\Delta r_\infty$  with that derived from an assumption about the magnitude of  $\mu d/\gamma$ .

## 62. THE GROWTH OF SURFACE COATINGS AND THIN FILMS FROM THE VAPOUR

In industrial practice, deposition from the vapour phase is often used to produce corrosion- or wear-resistant surface layers, or to enable engineering components to be used at higher temperatures. Even more significant, however, is the utilization of vapour growth in the electronics industry. Thin metallic or semiconducting films deposited on suitable substrates and with controlled dopants have very important technological applications, for example in the production of devices or integrated circuits. In recent years, the

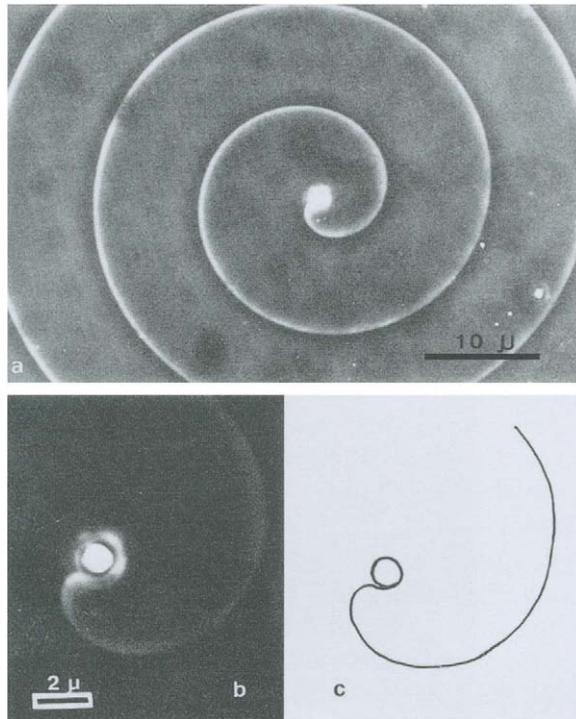


FIG. 13.15. (a) Growth spiral with step height 9 nm on hollow dislocation in SiC. (b) Enlargement of spiral centre to show reverse curvature as in (c).

importance of thin film technology has led to intensive study of various growth methods and structure-property relations.

There are several different techniques which allow carefully controlled growth from a vapour phase at realistic (albeit very slow) rates. The more important processes include physical vapour deposition (PVD), also called vacuum evaporation or vapour phase epitaxy (VPE), sputtering, chemical vapour deposition (CVD), molecular beam epitaxy (MBE) and chemical beam epitaxy (CBE). [Acronyms abound in descriptions of methods of crystal growth; the above examples are often further defined by additions such as metalorganic (MO) or organometallic (OM) (both are used) or “gas source” (GS) which give rise to descriptions such as MOCVD, MOMBE and GSMBE. When the product is amorphous or polycrystalline, the acronym usually ends in D for deposition, but when a single crystal is grown in a fixed orientation with respect to the orientation of the substrate, E is used to denote that the growth is, in some sense, “epitaxial”.] The methods differ in terms of the vapour source(s), which may be gaseous, liquid or solid, in the deposition, which may either represent condensation of the vapour or may involve a chemical reaction, and which may take place in a closed or in an open system, with or without a carrier gas.

In PVD, the vapour is usually obtained from a heated solid source or several sources (for compounds which do not evaporate congruently to give vapour of the same composition) and the reaction chamber is maintained at a pressure low enough to ensure that the mean free path is greater than the distance from source to specimen. Thus there are virtually no collisions of the atoms or molecules being deposited, and they travel in straight line trajectories from source to the depositing surface or elsewhere. The deposition thus has the characteristics of a beam, as distinct from the situation envisaged in Sections 60 and 61 where the vapour pressure exceeded the equilibrium vapour pressure. As congruent evaporation is rare, multiple sources are needed to grow intermetallic compounds or to introduce deliberate impurities ("dopants"), and rotation of the substrate combined with very close control of evaporation rates is then necessary to maintain composition homogeneity and uniform film thickness. An alternative to multiple sources is the use of very high local energy sources to ensure that all the atoms or molecules in a given locality are vaporized, i.e. to enforce congruent vaporization. This may be achieved by sputtering or by pulsed laser heating (LADA), in place of the more usual effusion cells which use electric, electron bombardment or induction heating. Problems of inhomogeneity also arise if any of the components are volatile (i.e. have a high vapour pressure at the deposition temperature). In such a case, it may be necessary to use a closed system which is sealed so that the total amount of each element is held constant, or a partly closed system in which the loss rate of the volatile constituents is minimized (Piper and Polich, 1961).

CVD involves either chemical reaction (e.g. reduction by hydrogen) or thermal decomposition of the vapour phase at the solid-vapour interface, and typically occurs at pressures between  $10^{-2}$  Torr and 1 Torr and at rather high temperatures. The oldest established process of this kind is probably the gas carburizing of steel, in which carbon is deposited on the (austenitic) surface as a result of the decomposition of either carbon monoxide or of a hydrocarbon such as methane. Other CVD surface treatments include nitriding from ammonia vapour and carbonitriding.

CVD may also be used to purify various metals by transport processes. Purification is achieved by reacting the impure element with a suitable gas to form a gaseous compound and then decomposing this compound at a different temperature where the pure metal is deposited and the released gas (in a closed system) is recirculated back to the impure source. In the Mond process for nickel, for example, the reactive gas is carbon monoxide, whereas in the van Arkel process for titanium, zirconium and other metals, it is iodine. In both of these examples, growth by decomposition of the gaseous compound takes place at a higher temperature than does dissolution by reaction, but in many other cases the growth temperature is lower than the evaporation temperature. Chemists frequently use simple closed systems containing chlorine or other halogens at a suitable pressure and with a temperature gradient to produce pure crystalline samples from initial "nutrient" material, but it should be noted that the product of such a transport process depends partly on the nucleation of the new crystals. It may, for example, lead to a cluster of small crystals, or to whisker growth, but it is unlikely to give a single crystal unless a seed crystal is introduced. When a seed crystal is used, bulk crystals (i.e. of millimetre

dimensions) may be produced by CVD methods but the growth times will generally be several days.

The chemical reactions used in closed systems must be reversible but this is unnecessary in open systems where gases flow through the reaction chamber and, in particular, the product gases of the decomposition which leads to growth are removed from the system. The decomposing gas itself may be formed by a reaction with a nutrient sample of the material to be grown, as in a closed system, or it may be supplied by an external source. The major use of open-system CVD deposition is in the growth of thin layers of semiconducting materials, and especially of silicon, by the decomposition of gases such as  $\text{SiCl}_4$ ,  $\text{SiHCl}_2$ ,  $\text{SiH}_2\text{Cl}_2$  and  $\text{SiH}_4$ .

Any description of the growth rate of an epitaxial single crystal thin film by CVD must clearly be more complex than that of a crystal growing from its own vapour. There are now a number of sequential processes required for growth; these include the decomposition reaction to produce the atoms or molecules for deposition, diffusion of these atoms or molecules across the viscous vapour boundary layer adjacent to the interface, and surface diffusion of ad-atoms to growth sites, as discussed in Sections 60 and 61. Because these various processes occur in series, each must proceed at the same rate (which is the actual growth rate) and this is achieved, as in all series processes, by suitably dividing the driving force among the various stages. It will often be the case that over a particular range of imposed conditions one process consumes the major portion of the driving force, so that the overall growth rate can be obtained to sufficient accuracy by considering this process alone and deriving its rate under the total driving force. This process is then said to "control" the overall rate. This can lead in general terms to either diffusion control or interface control of a growth process, or more specifically in the case of CVD to either of two diffusion controls or to chemical reaction rate control.

Figure 13.16 gives some experimental results for the CVD growth of silicon (Bloem and Classen, 1983). These results show that, at low temperatures of the substrate, the growth rate is negligible but, as the temperature rises, the growth rate increases in such a way that its logarithm is linearly proportional to  $T^{-1}$ . This fast linear increase continues up to about 1000°C (depending on the decomposing gas) but is then succeeded by a much slower rate of increase with increasing temperature. There are thus two different growth regimes: the rate of the chemical reaction controls the overall growth rate at relatively low temperatures of the substrate; and boundary layer diffusion controls it at higher temperatures. At the temperatures normally used for deposition (1200–1250°C), typical growth rates on silicon or sapphire substrates are about 2–4  $\text{nm s}^{-1}$ . The growth rate at room temperature is negligible in most CVD systems, so there is no risk of deposits building up on the cold walls of the reaction chamber.

Molecular beam epitaxy is the name given to a highly developed form of PVD which uses ultrahigh vacuum techniques and sophisticated sources, shutters and shields to give greater control over growth conditions. It is especially effective in growing films of intermetallic compounds and has been much used for group III/group V semiconductors such as GaAs and AlGaAs. A feature of MBE is that the atomic (e.g. Al, Ga or In) and molecular (e.g.  $\text{As}_4$  or  $\text{P}_4$ ) beams are all derived from effusion chamber sources in

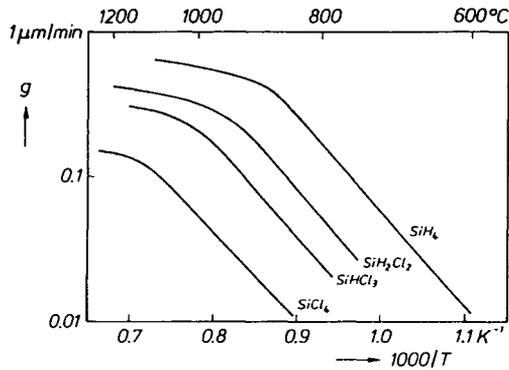


FIG. 13.16. Experimental variation of the CVD growth rate of silicon with temperature (after Bloem and Classen, 1983).

which the solid elements are heated to a suitable temperature. Such sources are difficult to control sufficiently finely to give the very stable and uniform reproducible growth required for many applications of thin film technology. Thus, although MBE has many successful applications and is now in widespread use with commercially available apparatus, a modified processing technique known as chemical beam deposition or chemical beam epitaxy (CBE) has recently been developed (Tsang, 1987; 1989). In CBE, the group III metals are added by the pyrolysis of suitable metalorganic compounds (e.g. trimethylindium or triethylgallium) which reach the heated substrate surface as molecular beams, and the group V metals are obtained from molecular beams of  $\text{As}_2$ ,  $\text{P}_2$  and  $\text{H}_2$  which are formed by the thermal decomposition, in a "cracker" at  $900\text{--}950^\circ\text{C}$ , of gaseous hydrides such as arsine and phosphine. Dopants are usually introduced as inorganometallic compounds and compounds of this type (disilane and germane) are also used for the CBE growth of group IV semiconductors. Conventional MBE thus utilizes heated solid sources whilst CBE uses sources which are all in the vapour phase at room temperature. Intermediate techniques are the so-called MOMBE and GSMBE in which the group III or group V element respectively is added from a chemical vapour source whilst the other element is added from a beam obtained by evaporation of a solid source.

The use of room temperature vapour sources in CBE has important advantages in controlling rates of flow, especially in compounds containing both As and P and in the growth of lattice matching layers of  $\text{Ga}_x\text{In}_{1-x}$  or  $\text{As}_{1-x}\text{P}_x$  on InP. As noted above, the individual beams for each element and dopant used in MBE are difficult to control sufficiently accurately but in CBE there is only one group III and one group V beam, because the various gas flows can be premixed.

MBE has been used to produce many sophisticated structures of new types; as one example, the strained layer heterostructures formed from alternating layers of Si and Ge or of Si and  $\text{Si}_{1-x}\text{Ge}_x$  should be mentioned. If the thickness of each alternate layer is constant, these are, in effect, synthetic, highly ordered structures of a new type and are

generally called superlattice structures (SLS). It is even possible, with suitable shutters, to produce single atomic layers by MBE.

We now consider the kinetics of deposition of a solid phase onto a relatively cold substrate by PVD or MBE or similar procedures, i.e. in circumstances where the growth does not depend on either chemical reactions or on diffusion through a boundary layer. In principle, the formation and growth of nuclei may be described in terms of the theory already developed but the high effective driving force implies that the critical nucleus size of classical theory may be as small as one or two atoms (or molecules) and, in these circumstances, macroscopic concepts clearly have to be abandoned and atomistic calculations are required. With the exception of very low deposition temperatures, it remains probable (Pound *et al.*, 1954) that nuclei (whether of single- or multilayer configuration) grow and shrink predominantly by the surface diffusion of ad-atoms rather than by direct accretion from the vapour. With slight modifications to allow for the difference in  $\Delta G'_c$  the rate of formation of nuclei will still be given by eqn. (60.32) provided a steady-state nucleation rate is achieved. Still using macroscopic concepts, it follows from eqns. (51.2) and (51.3) that a low value of the interfacial free energy  $\sigma^{\beta/S}$  between the deposited solid ( $\beta$ ) and the substrate ( $S$ ), which might be expected if there is an epitaxial relation, will help to reduce the critical free energy of formation of the  $\beta$  nucleus, but the value of the contact angle  $\theta$  is still determined largely by the ratio  $\sigma^{\beta/S}/\sigma^x$  of the surface free energies of the substrate and  $\beta$  phases in contact with the vapour. The critical supersaturation (or beam flux) required to achieve nucleation decreases as this ratio increases.

An epitaxial relationship between deposit and substrate may result from growth selectivity or from recrystallization of the deposited film, but it more usually indicates that nuclei of a particular orientation form more rapidly than any others. Such a preference may arise from a coherent  $\beta/S$  interface, or from a semi-coherent interface of the van der Merwe (1973) type (see Fig. 8.16), so that a reduction in  $\sigma^S$  is obtained at the expense of some coherency strain energy. In the special case that planes parallel to the interface in the stress-free growing phase and the substrate have similar structures, calculations indicate that a misfit of up to  $\sim 9\%$  may be accommodated completely by uniform elastic straining provided the film is extremely thin, i.e. contains only a few atomic layers. Matthews (1979) pointed out that such coherent films embody elastic strains three times larger than the maximum strains which have been measured in dislocation-free whiskers.

As the film thickness increases, it becomes energetically more favourable to reduce the long-range elastic field by formation of misfit dislocations, the field of which extends to about their separation (see Chapter 8). Various types of semi-coherent epitaxial interface and possible methods for their formation have been studied both experimentally and theoretically, and have been reviewed by Matthews (1979). The main results have been obtained from f.c.c. metals deposited on  $\{111\}$  substrates of other f.c.c. metals. The misfit dislocation model is not applicable to crystals with misfits much larger than 10%. However, epitaxy may also be obtained with incoherent interfaces if the appropriate surface free energies are lower for a particular orientation relationship. Pashley (1956) and Bassett *et al.* (1959) pointed out that epitaxial relations are frequently found for

isolated three-dimensional crystallites with no evidence of lattice strain. On this basis, they conclude that incoherent nucleation does not preclude an epitaxial orientation relation. Later work suggests that such favoured epitaxial orientations arise from the interphase analogues of coincidence site or other so-called "special" grain boundaries (see pp. 344–349).

At sufficiently low temperatures, the free energies of formation of nuclei of various orientations will be reduced to such an extent that random orientations are nucleated almost as readily as nuclei of the most favourable orientation. Thus there will be a temperature below which epitaxy caused by favourable nucleation is no longer observed, and this temperature will increase with increasing beam flux. In addition to the very small critical size, another important effect at low temperatures is the decrease in the surface mobility of the ad-atoms. If  $\tau^S$  [see eqn. (60.3)] becomes comparable with the experimental observation period, the ad-atom concentration must be time-dependent and a steady-state nucleation rate will not be achieved. Because each ad-atom has a negligible probability of re-evaporation, this condition is sometimes called "complete condensation". At very low temperatures, the time taken for an individual atom jump on the surface may be comparable with the time needed to accumulate a whole atom layer from the vapour. The concept of growth by accretion of ad-atoms is then inapplicable and the deposited film may be initially amorphous (see Section 63).

The atomistic theory of nucleation and growth of a crystalline phase on a substrate at temperatures where the ad-atoms retain some mobility will now be considered in more detail. This theory is generally considered to have originated in papers by Rhodin and Walton (Walton, 1962; Walton and Rhodin, 1963) but, despite some differences in formalism, many of the concepts and approximations which they used are equivalent to those of the classical theory. The atomistic description was further developed by Zinnsmeister (1968, 1969), Frankl and Venables (1970), Venables (1973), Stowell (1972, 1974) and others. It should be noted that the theory as it is developed in the following relates primarily to the "island" mode of growth (see p. 553), whereas that treated in Section 60 assumed layer growth. Island growth requires the nucleation of three-dimensional clusters on the substrate, which then continue to grow as distinct regions whereas, in layer growth, the nuclei are two-dimensional and may be few in number (of the order of one per atomic or molecular layer) because each spreads rapidly across the growing face. It follows that island growth will be favoured when atoms or molecules of the deposited phase bind more strongly to each other than to the substrate, and that the reverse situation applies to layer growth. Note, however, that in layer growth the influence of the substrate rapidly diminishes as successive layers are completed, and the structure will eventually reach that of the equilibrium bulk phase, with auto-epitaxial growth taking place. This implies that the binding energy decreases smoothly from its initial value (binding to the substrate) to its final value (self binding). However, if for some reason this decrease is not monotonic, so-called Stranski–Krastanov growth may be obtained. This might happen if the first layers are highly strained, or otherwise have a non-equilibrium structure, thus making further coherent deposition more difficult.

These distinctions are not absolute; for example, simultaneous multilayer growth, already discussed on p. 578, indicates a relatively high nucleation rate to growth rate ratio, i.e. a tendency towards island growth caused by a decrease in the relative strength of binding of an atom to the substrate and to existing deposits. Consider also the formation of the very first layer. When a fraction of this layer has been deposited, there may be few or many stable two-dimensional clusters on the substrate. In the latter case, the configuration might be described as (two-dimensional) island growth preceding layer growth.

The distinction between island and layer growth in terms of binding energies can also be expressed more macroscopically in terms of the surface and interfacial free energies. If  $\sigma^{vS} \geq \sigma^v + \sigma^S$ , there will be a negative net contribution to the free energy of formation of two-dimensional nuclei, and the only positive term will arise from the edge energy. This is a "wettability" condition; as stated on p. 453 (where the edge energy is not considered), the  $\beta$  phase will tend to spread indefinitely over the substrate surface. Clearly island growth is the more general assumption, and layer growth can be regarded as a special case; this is why the growth equations now to be developed appear more complex than those of Section 60. The shape of the three-dimensional nuclei in island growth is generally not known, and many calculations have assumed either hemispherical clusters or the slightly more general spherical cap shown in Fig. 10.9. Actually, this assumed shape is quite irrelevant to deposition under most beam conditions, where the critical nucleus size is extremely small.

Following Zinnsmeister, the time-dependent rate of change of clusters<sup>†</sup> of  $n$  atoms into those of  $n+1$  atoms is written in the form already given in eqn. (50.1), which is here supposed to apply to all values of  $n$  from two upwards. For nucleation on a substrate,  $Z_{n,t}$  of Section 50 must be replaced by the areal distribution function  ${}^S Z_{n,t}$  which specifies the number of clusters of size  $n$  per unit area of substrate at time  $t$ , and  $I_{n,t}$  similarly has to be replaced by  ${}^S I_{n,t}$ . Equation (50.1) becomes

$$d({}^S Z_{n,t})/dt = {}^S I_{n-1,t} - {}^S I_{n,t} \quad (62.1)$$

and the growth and shrinkage of clusters is regarded as continuous through dimers, trimers, etc., upwards. (The concept of the minimum number  $p$  of atoms needed to define a region of the new phase cannot be used here as the critical nucleus size of classical theory may be only two or three atoms.) Because the properties of very small clusters now influence the kinetics of nucleation, an additional factor to be considered is that these clusters may be mobile over the substrate. Although such mobility will probably decrease with increasing  $n$ , it may have various consequences; for example, stable nuclei may form by coalescence of subcritical nuclei in addition to the single atom growth process hitherto envisaged. However, the theory will first be developed on the assumption that only isolated ad-atoms are mobile, and the equations will later be modified to allow for cluster mobility. The rate of change of the ad-atom concentration is then given by the difference between the rate of arrival of new atoms from the beam flux,  $R$  (measured in atoms per

<sup>†</sup> Papers on the nucleation and growth of thin films from the vapour do not normally use the term "embryo".

unit area per unit time), and the rate of removal by evaporation or by absorption into any of the growing clusters. This gives:

$$d({}^S Z_{1,t})/dt = R - {}^S Z_{1,t}/\tau^S - 2{}^S I_{1,t} - \sum_{n=2}^{\infty} {}^S I_{n,t} \quad (62.2)$$

where  $\tau^S$  is given by eqn. (60.3) and the areal density of ad-atoms  ${}^S Z_{1,t} = \alpha^S/o$  provided  $\alpha^S$  is allowed, in contrast to eqn. (60.33), to vary with time.

A full description of the nucleation requires numerical solution of the above defining equations, but considerable simplification may be achieved by adopting the usual assumption of classical theory that there is some critical configuration of size  $n = n_c$  above which any cluster is stable (i.e. more likely to grow than to shrink) and below which any cluster is unstable (i.e. more likely to shrink than to grow). Experiments on thin film growth are concerned largely with measurements of parameters which combine nucleation and growth effects, such as the coverage of the substrate, the average size of a cluster, and the size and spatial distribution of visible clusters. Venables (1973) treated all the stable clusters of total areal density

$${}^S Z_{x,t} = \sum_{n_c+1}^{\infty} {}^S Z_{n,t} \quad (62.3)$$

as a group, so that eqn. (62.2) may be replaced by

$$d({}^S Z_{1,t})/dt = R - {}^S Z_{1,t}/\tau^S - 2{}^S I_{1,t} - \sum_{n=2}^{n_c} {}^S I_{n,t} - {}^S I_{x,t} \quad (62.4)$$

where  ${}^S I_{x,t}$  represents the loss of single atoms to stable clusters of average size  $n = x$ . A further simplification, again following the classical theory, is to assume that after an early stage of transient nucleation, during which the coverage of the substrate is still very small, a quasi-steady distribution of embryos for which  $n < n_c$  will be attained. A steady state requires eqn. (62.1) to be zero, so that each term on the right-hand side of eqn. (62.1) reduces to a constant, steady-state nucleation rate  ${}^S I$  for all  $n < n_c$ . Equations (62.3) and (62.4) then enable the density of stable clusters to be calculated as a function of time. The terms  $2{}^S I_1$  and  $\sum_{n < n_c} {}^S I_{n,t}$  in eqn. (62.4) are usually negligible so that the equation may be written

$$d({}^S Z_{1,t})/dt = R - {}^S Z_{1,t}/\tau - d(x{}^S Z_{x,t})/dt \quad (62.5)$$

The quasi-steady distribution for  $n < n_c$  must correspond approximately to the equilibrium distribution  ${}^S N_n$  of eqn. (60.34) but, in view of the very small size of the critical cluster, the difficulties discussed above in defining  $\Delta G_n$  in terms of macroscopic parameters are increased. Following Walton (1962), it is then more satisfactory to express the equilibrium between the clusters and the individual ad-atoms in terms of the energies  $E_n(s)$  required to dissociate clusters of size  $n$  and shape  $s$  into single ad-atoms. From

statistical mechanics or the principle of detailed balancing (see pp. 83 and 149), the distribution is specified by

$${}^S N_n = {}^S N_0 ({}^S N_1 / {}^S N_0)^n \sum_S C_n(s) \exp(E_n/kT) \quad (62.6)$$

where  ${}^S N_0$  is the number of ad-atom sites per unit area and the  $C_n(s)$  are statistical weights which depend on configurational contributions to the entropy of clusters specified by  $n$  and  $s$ . In most cases, only that shape  $s$  for which  $E_n(s)$  has a maximum value  $E_n$  makes a significant contribution, so that the sum may be replaced by the single term  $C_n \exp(E_n/kT)$ . Values of  $C_n$  for the ground states of a set of two-dimensional clusters were given by Frankl and Venables as  $C_n = 1, 3, 2, 3, 6, 6, 1$  for clusters with  $n = 1, 2, \dots, 7$  atoms on a substrate with a close-packed planar net of atoms; the numbers are all of order unity and the lowest values correspond to the most symmetrical clusters.

A quasi-steady-state nucleation rate  ${}^S I$  per unit area of substrate implies a constant rate of increase in the areal density of stable clusters, provided that each stable cluster, once formed, remains a separate entity in the assembly and that the fractional area of the substrate which is occupied by stable clusters is small. However, as clusters grow, there will eventually be some coalescence into larger units, thus reducing the overall density below the value predicted by the nucleation rate. Coalescence will result from impingement of clusters growing by accretion of single atoms, which is the assumption on which the above theory has been based but, as already mentioned, it is also necessary to consider the possibility that small stable clusters are themselves mobile and are thus able to move to, and join with, other clusters. The density of stable clusters may be increased by coalescence of mobile subcritical clusters and decreased by mobile supercritical clusters. Analysis indicates that the latter effect will generally be more significant, so that the rate of change of the density of stable clusters becomes

$$d({}^S Z_{s,t})/dt = {}^S I - {}^S I_{co,t} - {}^S I_{m,t} \quad (62.7)$$

where the last two terms represent the corrections due to coalescence by single atom growth and by the net effects of cluster mobility.

A further simplifying assumption, as in the Volmer theory of nucleation, is to neglect the fact that some stable clusters may shrink again and become unstable. The nucleation rate is then given by the product of the (quasi-stationary) density of critical clusters,  ${}^S Z_c$ , and the rate at which such a cluster of size  $n_c$  grows into a cluster of size  $n_c + 1$ . This gives

$${}^S I = {}^S Z_c (2\pi r_c r_1 {}^S Z_1) v \exp(-\epsilon^S/kT) \quad (62.8a)$$

$$= \beta_c^S Z_c {}^S Z_1 D^S \quad (62.8b)$$

$$= \gamma^S D^S ({}^S Z_1)^{n+1} \quad (62.8c)$$

where, in eqn. (62.8a), the simple assumption is made that the number of ad-atoms within one jump distance of a nucleus of critical size is  $2\pi r_c r_1 {}^S Z_1$ . Equation (62.8b) avoids this assumption by defining a "capture number"  $\beta_c^S$  for a critical cluster; the rate at which

individual ad-atoms join a stable cluster of size  $n > n_c$  can then be given by a similar equation with a different capture number  $\beta_n^S$ . The parameter  $\gamma^S$  in eqn. (62.8c) is defined from eqn. (62.6) as

$$\gamma^S = ({}^S N_0)^{1-n} \beta_c^S \exp(E_c/kT) \tag{62.9}$$

Note that the assumption of the Volmer theory may be avoided whilst retaining the equation in the form of eqns. (62.8b) or (62.8c) if a Zeldovich factor [see eqn. (60.35)] is incorporated into the definition of  $\beta_c^S$ .

The effect of cluster mobility on the nucleation rate is to add an additional term to eqn. (62.8) which may be written as a double sum

$$\sum_j \sum_k \beta_{jk}^S {}^S Z_{j,t} {}^S Z_{k,t} (D_j^S + D_k^S) \tag{62.10}$$

where the summation is over all values of  $j$  and  $k$  such that  $j - k < n_c$  and  $j + k > n_c$ ,  $D_j^S$  is the surface diffusion coefficient for a cluster of  $j$  atoms and  $\beta_{jk}^S$  is an appropriate capture number. Although the magnitudes of the cluster diffusion coefficients are not accurately known, it is expected that the sum [eqn. (62.10)] will generally represent a very small correction to eqn. (62.8). The term  ${}^S I_m$  of eqn. (62.7) is similarly given by

$${}^S I_m = 1/2 \sum_k \sum_l \beta_{kl}^S {}^S Z_{k,t} {}^S Z_{l,t} (D_k^S + D_l^S) \tag{62.11}$$

where both  $k$  and  $l$  in the summations are larger than  $n_c$ .

Coalescence by impingement due to atom-by-atom growth will become significant only when the density of stable clusters is sufficient to cover an appreciable fraction of the substrate. In approximate treatments of this difficult problem, the actual distribution of stable clusters of various size is replaced by the same total density  ${}^S Z_{x,t}$  of clusters of equal size  $x$  and surface area  $\sigma_x$ . The substrate coverage, i.e. the fractional area occupied by stable clusters, is then

$$C^S = {}^S Z_{x,t} \sigma_x \tag{62.12}$$

and various treatments of the coalescence (Vincent, 1971; Venables, 1973) give the rate of removal of clusters as

$${}^S I_{co,t} = 2 {}^S Z_{x,t} (dC^S/dt) \tag{62.13}$$

The last term of eqn. (62.5) replaces the corresponding term of eqn. (62.2) by the rate at which single atoms are absorbed by a set of stable clusters all of size  $x$ . This rate has three components representing respectively the atoms added by nucleation of new stable clusters, by direct impingement of beam atoms on to clusters of size  $x$  and by diffusion of single ad-atoms to these clusters. Hence

$$d(x {}^S Z_{x,t})/dt = (n_c + 1) {}^S I + (R\sigma_x + \beta_x^S {}^S Z_{1,t} D^S) {}^S Z_{x,t} \tag{62.14}$$

The nucleation contribution is almost always negligible, and the direct impingement term  $RC^S$  is important only at high coverage or at high temperatures when the diffusion term is limited by re-evaporation of ad-atoms. Equation (62.14) is more conveniently

expressed if the coverage  $C^S$  is used as the independent variable; its form then depends on the growth morphology of the clusters which influences the dependencies of  $x$  and  $o_x$  on the cluster dimension. If the average cluster spreads as a circular monolayer of radius  $r_x$ ,  $x$  and  $o_x$  both vary as  $r_x^2$ . As eqn. (62.14) gives the rate of change of the total number of atoms in stable clusters, the growth equation for two-dimensional islands may be written using eqn. (62.13) as

$$dC^S/dt = o(\beta_x^S D^S Z_{1,t}^S Z_{x,t} + RC^S) \quad (62.15)$$

whereas the corresponding equation for three-dimensional, hemispherical clusters is

$$dC^S/dt = (v/r_x)(\beta_x^S D^S Z_{1,t}^S Z_{x,t} + RC^S) \quad (62.16)$$

where  $v$  is the atomic volume. This equation may be expressed in a form which does not include  $r_x$ , as

$$[dC^S/dt] = [1 - d(\ln x)/3d(\ln C^S)] = v(\pi x/C^S)^{1/(1/2)}(\beta_x^S D^S Z_{1,t}^S Z_{x,t} + RC^S) \quad (62.17)$$

The remaining unknowns of this approximate analytic theory are the capture numbers. The capture number for a stable cluster depends upon the diffusive flow of atoms over the substrate to the cluster, so that a two-dimensional diffusion equation with appropriate boundary conditions must be solved in order to derive it. As discussed in Chapter 11, it is generally a sufficient approximation in such problems to assume steady-state diffusion, and solutions have been given for a "lattice" approximation (Stowell, 1972) and for a "uniform depletion" model (Lewis, 1970). The lattice model treats equi-sized clusters arranged on a two-dimensional close-packed net, and it is assumed that there is no ad-atom transfer between equivalent circular cells of this net. This procedure is similar to that used by Ham (1958) for three-dimensional growth (see Chapter 11). The effective capture number is found to be only weakly dependent on the radius of the cluster and, in terms of the substrate coverage  $C^S$ , may be expressed as

$$\beta_x^S = 4\pi/(\ln C^S) \quad (62.18)$$

to a sufficient approximation (Stowell, 1972). The uniform depletion model gives a similar dependence of the average capture number on  $C^S$  up to quite large values of  $C^S$  but the actual capture numbers are about one-half those given by eqn. (62.18). The analytic equations giving the capture numbers are generally complex, involving modified or unmodified Bessel functions, and depend upon limiting assumptions about the distribution of clusters (either randomly or on a lattice). These solutions are discussed in some detail by Venables (1973), Lewis and Rees (1974) and Lewis and Anderson (1978) but, as noted by Venables *et al.* (1984), it is generally adequate to consider the capture numbers to be slowly varying quantities in the range two to four for  $\beta_c^S$  and five to 10 for  $\beta_v^S$ , the average capture number for the stable clusters.

The description of the nucleation and growth of the deposited solid on the substrate is now given by eqns. (62.5), (62.7) and either (62.15) or (62.17) depending on the type of growth; the unknown quantities in these three coupled equations are specified in the various auxiliary equations already derived. A full treatment of the nucleation and growth

problem requires numerical solutions (see, e.g., Robertson and Pound, 1973), but some simple conclusions can nevertheless be reached from the form of the equations.

When a clean substrate surface is first exposed to the vapour, the single ad-atom concentration initially increases as  ${}^S Z_1 = Rt$  as only the first term in eqn. (62.5) is finite. At later stages, the ad-atom concentration begins to be limited by the other terms. Substituting from eqns. (62.12) and (62.14) into eqn. (62.5) gives

$$d({}^S Z_{1,t})/dt = R(1 - C^S) - ({}^S Z_{1,t})/\tau \quad (62.19)$$

where

$$\tau^{-1} = (\tau^S)^{-1} + \beta_N^S D^S {}^S Z_{N,t} \quad (62.20)$$

${}^S Z_{1,t}$  increases in the initial stage of transient nucleation until the quasi-steady distribution of subcritical clusters is established; this transient range extends approximately to a time  $t = \tau$  given by eqn. (62.20), after which the right-hand side of eqn. (62.19) is effectively zero and  ${}^S Z_1$  is constant. The number of stable clusters which form during transient nucleation may now be estimated by integrating eqn. (62.7) from  $t=0$  to  $t=\tau$  (neglecting the corrections for coalescence and cluster mobility) and using eqn. (62.8c) with  ${}^S Z_1 = R\tau$ . This gives

$${}^S Z_{N,t} = \gamma^S D^S R^{n+1} \tau^{n+2} \quad (62.21)$$

Equations (62.20) and (62.21) may be solved for  ${}^S Z_{N,t}$  and there are simple analytical formulations for the limiting cases where  $\beta_N^S D^S {}^S Z_{N,t} \tau^S$  is much smaller or much greater than unity, i.e. where  $\tau$  is approximately equal to or is much smaller than  $\tau^S$ . As  $D^S \tau^S = (\lambda^S)^2$ , which decreases with increasing temperature [see eqn. (60.4)], these two limits correspond to high and low temperature conditions. At high temperatures, transient nucleation ends at a time  $t \leq \tau^S$ , and the density of stable clusters formed during the transient period is negligible. At low temperature, however, the density of stable clusters formed during the transient period

$${}^S Z_{N,t} = \left[ \frac{\gamma^S (R/D^S)^{n+1}}{(n_c + 2)(\beta_N^S)^{n+2}} \right]^{1/(n+3)} \quad (62.22)$$

may be significant. Before the expression can be evaluated, it is necessary to assume a value for  $\beta_N^S$ . Stowell (1972) suggested that the appropriate coverage for use in eqn. (62.18) is given approximately by the assumption that at this stage each stable cluster is only just supercritical and thus contains  $n_c + 1$  atoms. Computations by Frankl and Venables (1970) support this assumption of little growth during the period of transient nucleation.

The condition for some stable nuclei to form during the transient stage  $0 \leq t \leq \tau$  was given above as

$$A_N = \beta_N^S D^S \tau^S {}^S Z_N \gg 1 \quad (62.23a)$$

and is called "complete condensation", because all atoms arriving at the substrate are assumed to remain in the deposit. Frankl and Venables showed that, if  $\beta_N^S$  is assumed to

vary as  $x^{1/3}$ , the density of stable clusters then saturates at time  $\tau$ , but that with the type of dependence given in eqn. (62.18),  ${}^S Z_{x,t}$  continues to increase slowly until either coalescence or cluster mobility produces a limiting density. The other (high temperature) limit was called by Stowell and Hutchinson (1971) "extreme incomplete condensation", as most of the ad-atoms are not captured by stable clusters but re-evaporate, and the first term in eqn. (62.14) (representing direct impingement) is the only significant contribution to cluster growth. The condition for extreme incomplete condensation is

$$A_x \ll C^S \quad (62.23b)$$

An intermediate regime, called "initially incomplete condensation" arises when clusters capture some ad-atoms by surface diffusion, so that the second term in eqn. (62.14) is the more significant; however, in the early stages, at least, some ad-atoms re-evaporate. When the coverage is sufficiently extensive to cause appreciable coalescence, however, virtually all atoms from the beam may be captured, so this type of condensation later becomes complete.

From eqn. (62.19), it follows that near saturation and indeed for all times  $t \geq \tau$ , the density of single ad-atoms is given by

$${}^S Z_1 = R(1 - C^S)\tau^S / (1 + A_x) \quad (62.24)$$

and is thus effectively constant. The rate equation (62.7) in conjunction with eqns. (62.8c), (62.13) and (62.23) becomes

$$d({}^S Z_{x,t})/dt = (1 - C^S)\gamma^S D^S (R\tau)^{n+1} - 2^S Z_{x,t}(dC^S/dt) - {}^S I_m \quad (62.25)$$

This equation shows that at high temperatures where  $\tau = \tau^S$  and the time interval of transient nucleation is negligible, the initial (steady-state) nucleation rate for different fluxes varies as the  $(n_c + 1)$ th power of  $R$ . Measurements have been made of the nucleation rate for deposition of noble metals (gold and silver) on cleaved (100) substrate of several different alkali halides. The gold or silver forms island deposits, the density and size distribution of which after a flux  $R$  for a time  $t$  at a temperature  $T$  may be measured by "fixing" the structure with an evaporated thin carbon film, dissolving away the substrate and examining the carbon foil with its attached island clusters in the electron microscope. (Note that this procedure assumes that all stable clusters remain unchanged after the beam is switched off and during the subsequent preparation of the specimens for electron microscopy. This is generally true for island growth, but not for layer growth, where the deposit may disappear again when the beam is removed.) All measurements of the high temperature nucleation rate of metals on alkali halides indicate a dependence of  ${}^S I$  on  $R^2$ , which has been interpreted to mean that a single ad-atom constitutes a critical nucleus for growth in these systems.

At later times,  ${}^S Z_x$  is limited by growth coalescence of either immobile or mobile clusters. If the mobility term  ${}^S I_m$  of eqn. (62.7) is ignored, Stowell (1972, 1974) showed that the maximum density of stable clusters may be expressed in terms of the coverage of the substrate at which the maximum is observed. The relation depends on whether

three-dimensional or two-dimensional islands are nucleated; for three-dimensional islands, it may be written in the form

$$({}^S Z_{v,t}/{}^S N_0)^{3/2}(C^S + A_{v,t})(1 + A_{v,t})^n = ({}^S N_0)^{1-n}(R/D^S)^n(D^S \tau^S)^n \exp(-E_c/kT) \quad (62.26a)$$

whereas the corresponding result for two-dimensional islands is simply

$$({}^S Z_{v,t}/{}^S N_0)(C^S + A_{v,t})(1 + A_{v,t})^n = ({}^S N_0)^{1-n} R^n D^S (\tau^S)^{n+1} \exp(-E_c/kT) \quad (62.26b)$$

where, in both cases, the coverage and the density of stable clusters are evaluated at the maximum (saturation) value of the latter. This maximum in  ${}^S Z_{v,t}$  occurs at quite low values of  $C^S$  (typically  $< 0.2$ ) in all cases and may readily be determined experimentally even at temperatures where the nucleation rate is too rapid to be measurable. From eqn. (62.26a), it follows that the maximum cluster density varies as  $R^p$ , where  $p = 2n_c/3$  for extreme incomplete condensation forming three-dimensional islands at high temperatures and  $p = n_c/(2.5 + n_c)$  for complete condensation forming similar islands at low temperatures. The corresponding values of  $p$  for two-dimensional islands at high and low temperatures respectively are  $n_c$  and  $n_c/(n_c + 2)$ .

The evolution of the single atom density  ${}^S Z_1$  and of the stable cluster density  ${}^S Z_v$  with time for high and low temperatures of deposition is shown schematically in Fig. 13.11, which is taken from the review paper by Venables *et al.* (1983). At high temperatures,  ${}^S Z_{v,t}$  (and hence  $C^S$ ) are not appreciable until some time after the steady nucleation rate has been achieved at  $t = \tau^S$ . For times greater than  $\tau^S$ , the ad-atom density  ${}^S Z_{1,t}$  remains constant at the values  $R\tau^S$  whilst the density  ${}^S Z_{v,t}$  of stable clusters steadily increases. Eventually, however, coalescence causes both densities to decrease again. The single ad-atom density increases for a much longer time and reaches a much higher value at low temperatures where re-evaporation is unimportant. The stable cluster density continues to increase beyond  $t = \tau$  [eqn. (62.20)] at which time the single ad-atom density reaches a maximum; most of the nucleation occurs near this peak.

As already indicated, experimental measurements of the number, mean size and size distribution of stable clusters may readily be made by electron microscopy, and other techniques give the total amount of material deposited and hence derived quantities such as the incremental and total condensation coefficients. These experimental results have been used in various tests of the above theory; in particular, combination of the value of  $n_c$  obtained from the nucleation rate measured at high temperatures with the dependence of the maximum cluster density on  $R$  and  $T$  has enabled a consistent interpretation over large temperature ranges to be given in many cases. From an Arrhenius plot of the temperature dependence, an experimental activation energy  $E$  may be obtained where it follows from eqn. (62.26a) that for three-dimensional island growth

$$E = 2[E_c + (n_c + 1)w^S - \varepsilon^S]/3 \quad (62.27a)$$

and

$$E = (E_c + n_c \varepsilon^S)/(n_c + 2.5) \quad (62.27b)$$

for extreme incomplete and complete condensation respectively. The corresponding activation energies for incomplete and complete two-dimensional islands are

$$E = E_c + (n_c + 1)w^S - \varepsilon^S \quad (62.27c)$$

and

$$E = (E_c + n_c \varepsilon^S)/(n_c + 2) \quad (62.27d)$$

Stowell (1972) analysed the results of Robinson and Robins (1970) for the growth of gold deposits on potassium chloride substrates and deduced values of  $\nu$ ,  $\varepsilon^S$ ,  $D_0$  and  $w^S$  which are reasonable and self-consistent over wide temperature ranges. Many similar analyses of other noble metal-alkali halide data which were made subsequently (see Venables *et al.*, 1984) also give self-consistent values of the activation energies for evaporation and surface diffusion respectively of an ad-atom. However, later more detailed comparison of experimental results with the theory revealed some discrepancies; for example, the overestimation of the value of the coverage  $C^S$  and the value of  $n_N$  at the maximum in  $n_N$  at high temperatures. This effect was attributed to the neglect of cluster mobility and was demonstrated by Stowell (1974) by means of a plot of  $\log(x^S)^2 = \log(D^S \tau^S)$  versus  $1/T$ , which is linear as expected over most of the temperature range, but which shows an anomalous increase in  $x^S$  at high temperatures. Direct measurements of  $\tau^S$  (Knabbe and Harsdorff, 1979; Harsdorff and Knabbe, 1979), although very difficult because of the low value of  ${}^S Z_{1/N_0} (< 5 \times 10^{-8}$  for Au on NaCl) at which nucleation occurs, indicated nevertheless that the value of  $\varepsilon^S$  deduced from electron microscopic observations on Au films on NaCl substrates is too high by about 0.1–0.2 eV, and measurements of the integrated condensation coefficient (Reichelt *et al.*, 1980) strongly suggest that the difference in energy  $\varepsilon^S - w^S$  for the same system is considerably lower than the value deduced from cluster measurements. A probable source of error in the early cluster measurements was discovered by Henry *et al.* (1976) and by Chapon *et al.* (1976), who found that some of the smaller stable clusters were not revealed by the carbon film technique, but became visible if the Au clusters were decorated with Cd before the electron microscopic examination. Higher values of  ${}^S Z_{N,I}$  were found by means of the Cd decoration method at short times, and in the case of Au/NaCl there was also an anomalous dip in  ${}^S Z_{N,I}$  at very short times; this is also thought to be an effect of cluster mobility (Chapon and Henry, 1981).

Other experimental results, e.g. Krohn and Bethge (1979) and Velfe *et al.* (1982), show that impurities have very marked effects on the maximum value of  ${}^S Z_{N,I}$ , which in the case of very pure NaCl substrates is only about 10% of that found in the earlier measurements. The later work with the very pure substrates gives further evidence for cluster-mobility effects and also (Knabbe and Harsdorff, 1979; Velfe and Krohn, 1982) for the possibility that dimers, trimers, etc. may be directly deposited from the vapour source. Further complications arise from nucleation at defect sites (Trofimov, 1981; Usher and Robins, 1982) which may be continuously created, e.g. by stray electrons emitted from the metal vapour source. Thus if the number of such sites is proportional to the flux  $R$ , the experimental result that  ${}^S I$  varies as  $R^2$  may indicate heterogeneous nucleation on

defects rather than homogeneous nucleation with  $n_c \cong 1$ . However, by suitably modifying the gold source, Usher and Robins were able to eliminate this beam-induced defect production and their results then showed cluster mobility effects at and above 133 K. The kinetic analysis indicated that at 128 K and lower temperatures, only single ad-atoms were mobile.

There are thus many complicating features to be taken into account in the analysis of island growth in any real system, and this means that a large number of (often unknown) parameters may have to be specified; several such analyses have nevertheless been made. Despite difficulties with the details, the theory outlined above gives a generally satisfactory description of the experimental observations, so that it may be claimed that island growth is reasonably well understood. Now consider again the other extreme, i.e. layer growth.

From the outline description on p. 561, it is evident that layer growth may arise either from pre-existing steps or from two-dimensional (disc-shaped) island nuclei, one atomic or molecular layer in thickness. The essential feature of layer growth is that once a step (either closed or self-perpetuating) is formed, it can spread rapidly over the preceding layer, covering it with a new layer if the crystal is growing, or exposing it to the vapour if the crystal is evaporating. Models for layer growth (or evaporation) when the super- (or under-) saturation is small have been discussed extensively in Sections 60 and 61, and it remains only to consider layer growth under beam conditions of high effective supersaturation. Many metal films grow on metal substrates in this way provided the mismatch between the two lattices across the interface is not large; in the most favourable case of auto-epitaxy, when growth on a substrate is equivalent to growth of a bulk single crystal, only layer type growth is to be expected. Many important examples of semiconductor films are also produced by layer growth on substrates.

Under beam conditions, each layer may arise from many nuclei, and new layers may be nucleated on partly completed layers, giving so-called simultaneous multilayer growth, discussed on p. 578. As the supersaturation increases, so does the number of incomplete layers at the growth interface, and this corresponds to a kinetically induced surface roughening transition. An accurate description of such growth requires computer simulation, as mentioned on p. 562, but approximate descriptions may be based on the modified kinetic theory developed in this section.

Experimental observations of layer growth are also intrinsically very difficult so that not very many layer growth systems have been studied in as much detail as island growth systems. Electron microscopic identifications of nucleation sites, or measurements of the densities of nuclei, for example, are possible only if techniques exist for making steps of atomic height visible. One such technique is the gold decoration method used for alkali halide crystals which depends on the formation of three-dimensional island nuclei of gold at the steps. Experimental work with this technique has been concerned both with the formation of two-dimensional "island" nuclei during growth and two-dimensional "hole nuclei" ("lochkeime") during evaporation. Early results on the layer growth of potassium chloride showed that two-dimensional island nuclei were formed on the flat terraces between pre-existing steps at critical supersaturations exceeding about three at 330°C and

$\sim 1.6$  at  $400^\circ\text{C}$ , and lochkeime were formed below critical undersaturations approximately reciprocal to these values. In later work with very pure KCl, however, Stein and Meyer (1980) observed that nucleation occurs only for supersaturations exceeding about seven and lochkeime densities during evaporation were reduced by up to five times. This strongly suggests that much of the nucleation in the early work was heterogeneous at special sites associated with impurities; we have already noted that some lochkeime are catalysed by dislocations.

Later measurements of the nucleation rate  $^SI$  and of  $^SZ_{v,i}$  in the ranges  $289\text{--}343^\circ\text{C}$  and  $10\text{--}40$  for the substrate temperature and supersaturation ratio respectively were made by Meyer and Stein (1980, 1982). A typical nucleation rate is  $^SI \cong 6 \times 10^{11} \text{ m}^{-2} \text{ s}^{-1}$  at  $589 \text{ K}$  and a supersaturation ratio of 17. The nucleation rate is represented by a modified form of eqn. (60.32) and this leads to a value for the edge energy [eqn. (60.20)] of  $\sim 7.5 \times 10^{-11} \text{ J m}^{-1} = 0.075 \text{ nJ m}^{-1}$ , which is of the expected order of magnitude. The corresponding critical nucleus size  $n_c$  varied from four to 10 "molecules" in the temperature and supersaturation ranges investigated. Laaser *et al.* (1983) modified these figures slightly in a more sophisticated interpretation of later experimental work which produced a critical nucleus size of only one molecule for evaporation at sufficiently large undersaturation.

There are many experimental results on the layer growth of f.c.c. metals on substrates of other f.c.c. metals. In the earlier work, only noble metals plus copper were used, but later results concern more reactive metals like iron and nickel. Interest has centred largely on the epitaxial relations between the substrate and the deposit, and there are relatively few observations of the mechanism or kinetics of the deposition. Some electron microscopic studies of the formation of two-dimensional island nuclei have been made by depositing less than a monolayer of palladium or lead on an atomically flat  $\{111\}$  silver substrate. Yagi *et al.* (1979) were able to estimate palladium cluster densities of  $\sim 3 \times 10^{14} \text{ m}^{-2}$  at  $150^\circ\text{C}$ ; similar results for lead deposited on silver indicate a smaller density of nuclei and this suggests that the binding energies  $E_i$  of lead atoms in a two-dimensional cluster are smaller than those of palladium atoms.

Finally, many metal-metal systems have been found to grow in the intermediate Stranksi-Krastanov mode (see p. 553); cases which have been studied in detail by electron microscopic and other techniques include the deposition of silver on  $\{110\}$  tungsten and  $\{100\}$  molybdenum substrates. The results show that, once islands have started to form, the island density  $Z_{v,i}$  is constant over wide ranges of deposition (e.g. from three to 150 monolayers in the case of the  $\{110\}$  tungsten substrate), but  $Z_v$  has a larger variation with the substrate temperature than in the case of island growth. The results available for the growth of silver on tungsten, molybdenum and silicon were analysed by Spiller *et al.* (1983) and by Venables *et al.* (1984), who concluded that the critical nucleus size was seven to 20 atoms in all cases, in contrast to the single atom critical size of many island growth systems. This analysis assumes that an intermediate absorbed layer (or several layers) forms first and that the island nuclei are two-dimensional. The larger critical nucleus size may be attributed to the fact that, in this mode, island growth is only very slightly preferred to layer growth.

### 63. GROWTH OF WHISKERS, NANOCRYSTALS, QUASI-CRYSTALS AND AMORPHOUS SOLIDS FROM THE VAPOUR PHASE

In this section, some special conditions of growth from the vapour phase which may lead to crystals with one or more very small dimensions, or to solids without the translational symmetry which characterizes a crystal, are examined. A three-dimensional crystal can be stable only if it is larger than a critical nucleus, and a crystal face can grow by the dislocation mechanism only if the separation of adjacent opposite dislocations exceeds  $2\rho_c$ . These considerations provide a theoretical lower limit to the dimensions of a crystal grown at a fixed supersaturation and, if the vapour pressure is drastically reduced, following a growth process, the smallest crystals should, in principle, re-evaporate and then, in a closed system, re-condense on the largest crystals. This is the vapour phase equivalent of "Ostwald ripening", as originally described for solids growing from liquid solution, or of "coarsening" in two-phase solids. At sufficiently low temperatures, however, the rates of evaporation and condensation become negligible, so that crystals of small dimensions may remain metastable indefinitely.

Solids may be "small" in one, two or three dimensions, giving rise to growth morphologies which are respectively fibrous, plate-like or equi-axed. Fibrous (or whisker) growth is observed to occur naturally in many minerals, organic compounds, polymers, etc. and may often be attributed to particular features of the crystal structure. Nabarro and Jackson (1958) discuss, for example, aligned molecular "chains" (as in cellulose and fibrous silicates), molecular layers curled to form hollow "scrolls" [as in chrysotile (asbestos) and some paraffin waxes] and planar molecules arranged in columns with strong electrostatic forces between adjacent molecules in one column, but weak forces between adjacent columns (as in metaldehyde and trioxane). We shall consider only metal whiskers, many of which have cubic structures, thus making it clear that the observed fibrous growth cannot be attributed to anisotropic molecular or atomic forces. When there is no special axis, uniaxial growth can only be caused by some asymmetry in the imposed growth conditions, or in the perfection of the crystal structure or the free surface of the growing crystal.

Naturally occurring "hair silver" was probably the first metal whisker to be noticed and a procedure for its artificial manufacture was published in the middle of the sixteenth century. In modern times, fibres, and more especially "whiskers", first attracted attention because of their potential technological advantages in composite materials. More recently, interest in the possible applications of ultrafine particles has increased, and there is now a rapidly developing "nanoscience" and "nanotechnology" devoted to studies of the properties and applications of aggregates of such particles (Gleiter, 1989; Gutmanas, 1990; Uyeda, 1991).

A whisker is a fibre which is a single crystal, and whiskers may be grown by many different techniques, including deposition from the vapour or from liquid solution, decomposition of a chemical compound, electrolytic deposition and growth entirely from the solid. Following Hardy (1956) and Nabarro and Jackson (1958), a distinction is made between vapour or liquid phase growth and the growth of whiskers from the surface of a solid. Whiskers grown from the solid are described as "proper" whiskers or, if grown more

rapidly under the influence of an applied mechanical stress, as "squeeze" whiskers. The validity of the distinction is emphasized by experimental results which show that a whisker grows from the vapour or liquid phases by addition of atoms at its tip whereas a solid phase whisker grows from its base. Many whiskers have diameters of the order of 1–10  $\mu\text{m}$ , but mercury whiskers of diameters of only 10–15 nm and up to 1 mm in length were prepared in the early experiments of Sears (1955). The theory of dislocation-assisted growth from the vapour or liquid phases naturally leads to the hypothesis that whiskers have a single axial dislocation of screw or near-screw type, thus accounting for uniaxial growth in a crystal structure which does not possess a unique axis.

A large number of experiments on the growth of whiskers of different metals from the vapour phase have shown that it takes place most readily on a substrate of the massive metal and in the presence of an inert gas. Coleman and Sears (1957) found, for example, that very pure zinc whiskers of 1–3  $\mu\text{m}$  in diameter and up to 17 mm in length were obtained from the vapour in atmospheres of hydrogen or helium at pressures from  $1.3 \times 10^3$  to  $8 \times 10^4$  Pa, but only massive zinc was deposited at lower pressures. The diameters of the whiskers decreased to a few tenths of a micron as the pressure increased to the upper limiting value. Evidence that atoms are deposited at the tip of a zinc whisker had previously been found by Sears and Coleman (1956), who observed that growth ceased immediately a whisker accidentally encountered another whisker. At a later stage of the growth, however, whiskers are frequently observed to thicken at their bases. The growth of a whisker eventually ceases, and this occurs sooner and at a shorter whisker length if the carrier gas contains impurities which are absorbed on the whisker surface.

In terms of the descriptions already given, the absence of sideways growth of a whisker may be explained if the sides are low index planes without emergent dislocations, so that growth is impossible at low degrees of supersaturation. Many experimental results are consistent with this hypothesis, the growth direction generally being a low index direction and the side faces low index planes. There are two possible reasons for the continued growth at the tip. One is that an axial dislocation line allows the deposition of atoms at the tip by the Frank mechanism. The other is that there is some diffusion control of the growth, and that the supersaturation of the vapour near the tip is larger than it is over the side surfaces. It is improbable that this "diffusion point effect" could be large enough to stabilize whisker growth, but if sideways growth is temporarily stopped for any reason, the whisker morphology may be stabilized by the poisoning of the side surfaces by impurities.

Evidence in favour of the dislocation mechanism is unconvincing: some whiskers appear to be perfect crystals, others to contain a single axial dislocation as required by the theory, and still others to contain many dislocations. It is not easy to verify the presence of an axial screw dislocation; X-ray measurements of the consequent lattice twist by means of a high resolution Laue technique indicated the net axial Burgers vector to be zero in a number of whiskers of different metals (Webb, 1958). The sole exception was palladium, in which the net Burgers vector was one to 10 times the basic (shortest) lattice repeat vector. Gomer (1957) used an ingenious method to verify a twist in some of his mercury whiskers by applying an electrostatic stress to a specimen under observation in a field ion

microscope. The fact that most metal whiskers (in contrast to alkali halide whiskers) do not exhibit the predicted twist does not necessarily invalidate the dislocation growth mechanism, as the dislocation responsible for the growth may have subsequently climbed or glided out of the crystal, thus bringing growth to an end. Alternatively, a dipole of two opposite dislocations would produce zero twist whilst still (as discussed in Section 61) promoting growth. However, in the absence of positive evidence for dislocation growth, the surface-poisoning mechanism must also be considered.

In discussing the clustering of steps on p. 567, it was pointed out that an impurity which reaches adsorption equilibrium relatively slowly will stabilize bunches of steps, as it will be more effective on flat regions of surface which have been exposed to its effect for some considerable time. The same concept applies even more directly to the growth of a crystal as a whole, as surfaces which are stationary may become covered with dirt which effectively prevents further growth. Thus a temporary fluctuation in growth rate may lead to a permanent inhibition of growth, whilst surfaces which keep growing remain clean. In the case of a nearly perfect crystal without any dislocations, the adsorbed impurities might impede either the nucleation or the growth of new layers, or both, and once a needle-like shape has been established, uniaxial growth could persist. However, the nature of the fluctuation which starts this process is not clear.

When whiskers grow at the tip either by the dislocation or impurity mechanisms, the growth rate is much larger than that calculated from the atom flux from the vapour to the tip. The whisker thus grows by acquiring atoms which are first adsorbed on the sides and then diffuse over the surface to the tip. The tip may be regarded as a sink, and the growth rate will increase with the length of the whisker until this length exceeds the surface diffusion distance, after which approximately 50% of the atoms deposited on the side surfaces within this diffusion distance of the tip are eventually captured by the tip. Equally, about one-half of the atoms adsorbed within a sufficient distance of the base of the whisker will reach this base before they re-evaporate. When the whisker has nucleated on a substrate of the same metal, this will provide a sink for these diffusing atoms, but if the substrate is, for example, glass, an appreciable concentration of adsorbed atoms may be formed around the base. The re-entrant angle at the base may then provide a favourable site for the nucleation of new layers on the side surfaces, so that the whisker may begin to thicken quite early in its growth. Such an effect has been observed in the growth of zinc whiskers on glass.

The theory of whisker growth from solution or by the decomposition of a chemical compound in the vapour phase is similar to that of growth from its own vapour, and need not be considered further. The hydrogen reduction of metallic halides is a commonly used method for the preparation of metallic whiskers. Conditions similar to those giving whiskers often result in the growth of small platelets, i.e. to crystals with two rapid growth directions. Some other very striking growth morphologies have been observed, including whiskers with sharp kinks or whiskers bent into continuous helices. It appears that these complex forms are single crystals (i.e. true whiskers) in most cases, but they are frequently not perfect. As a general rule, the larger the whisker, the greater is the number of imperfections it contains.

In contrast to whiskers and platelets, ultrafine particles have no macroscopic dimension; they can be loosely defined as particles of submicroscopic size, resolvable only in an electron microscope (Uyeda, 1991). Noting that this definition covers a size range from a single atom or molecule to a few microns, Uyeda further proposed that the term "cluster" should be used for particles at the lower end of the size range, and "microcluster" for particles containing a countable number of atoms or molecules. There is, however, no general agreement on nomenclature in this new and rapidly developing area; Gleiter refers to nanometer-sized clusters and Gutmanas to ultrafine powders. Polycrystalline aggregates of ultrafine particles, whether obtained by processing of fine powders or by other methods, are called nanocrystalline materials by Gleiter, and it seems likely that this term will gain wide acceptance.

Ultrafine particles may be obtained by various techniques; those of interest here are related to vapour phase aggregation and condensation, but others include formation of very fine liquid droplets, formation of colloidal suspensions and other methods of precipitation from solution, spark erosion and ball milling of powders, etc. A mixture of ionized and neutral monomers, dimers and higher clusters may be obtained directly by forced emission from a surface by high energy methods such as sputtering or laser irradiation. Although useful for certain special applications, such methods are not efficient in producing clusters as most of the ejected atoms are monomers, dimers, etc. and there are relatively few atoms in larger clusters.

One method for producing individual very small particles (10 nm or less) has already been discussed, namely evaporation in a vacuum on to a substrate under conditions of island growth. The morphology of each particle may be studied and will be related to the surface free energies. The particles are, however, not necessarily single crystals; for noble metals evaporated on to a NaCl substrate, for example, Ino (1966) found the particles to be multiply twinned, with either five or 20 tetrahedral crystallites of the expected f.c.c. structure, each in a twin orientation relative to its neighbours. The external shape is a pentagonal decahedron for five units and an icosahedron for 20 units.

The disadvantage of the island growth method is that it is difficult to collect the small particles which are dispersed over the substrate surface. A modified procedure called "VEROS" (Vacuum Evaporation on a Running Oil Surface) was introduced by Yatsuya *et al.* (1978) to overcome this difficulty. An oil film on the underside of a rotating disc above the vapour source acts as the deposition surface. The oil spreads outwards from the centre of the rotating disc, carrying with it any clusters formed near the centre which grow into ultrafine particles before reaching the periphery of the disc. These particles are collected with the oil as it leaves the disc, and are separated from it by vacuum distillation. Particles of 3–8 nm in a wide range of metals have been produced by this method, but the particle size usually increases during the subsequent distillation. Modifications of the VEROS method have been described by Nakatani *et al.* (1987) and by Hayashi (see Uyeda, 1991). Hayashi used a disc cooled by liquid nitrogen and, instead of the oil film, he evaporated an organic solvent simultaneously with the metal, so that the metal particles are embedded in the solidified solvent.

The most important physical technique for producing ultrafine particles has been variously described as gas evaporation and gas condensation; we use the combined name since both processes are involved. A vapour of monomers is obtained by evaporation into an inert gas at a pressure of  $10^2 - 10^3$  Pa, and collisions of the vapour atoms or molecules with those of the gas effectively "cool" the vapour and lead to the nucleation of small clusters which can then grow by collisions with monomers and with other clusters. The clusters reach a limiting size which depends on the evaporation rate and on the atomic weight of the inert gas. The vapour is produced most simply by evaporation from a heated boat or crucible containing molten metal or, in suitable cases, by sublimation from solid metal, and the mean particle size may be controlled by varying the evaporation rate and the pressure of the inert gas. Ceramic oxides, hydrides, nitrides, etc. may be obtained in ultrafine particle form by first producing metal clusters and then admitting oxygen, hydrogen, nitrogen or ammonia into the reaction chamber. An alternative technique is to use lasers to produce a vapour of (mainly neutral) atoms by evaporation from the surface of a target in a suitable inert gas environment. This "laser ablation" technique is very versatile; in particular, it is not limited to metals but may be used for many alloys or compounds.

Other modifications of the gas evaporation–condensation technique include plasma flame, electron beam, arc discharge or high frequency induction heating. The various evaporation methods all produce relatively pure particles with clean surfaces and few lattice defects, but they are expensive to operate, and much cheaper bulk production is usually possible by methods involving chemical reactions. In the gas evaporation–condensation technique, the particles are first apparent as a plume of "metal smoke". An analysis of a smoke of candle-flame shape (see Uyeda, 1991 for detailed description and references) shows that there are three zones. In the inner zone, the particles are small and isometric, in the intermediate zone, they are much larger but still isometric, whilst in the outer zone, platelets or rods may be formed. There is also a particle-free, vapour zone immediately adjacent to the vapour source. Nucleation is believed to occur at the limit of the vapour zone and growth in the intermediate zone is believed to depend on diffusion of vapour through the inner zone. The outer zone may be formed from vapour diffused below the source and then blown upwards by convection. Clearly, the conditions in such a smoke are quite complex, and any detailed quantitative treatment is very difficult. The candle-flame smoke has been extensively investigated by Uyeda and his group at Nagoya; it seems to arise when the evaporation is carried out from a boat or a spiral heater. If a crucible is used, there is only a single zone (Grandqvist and Buhrman, 1976).

The structures and morphologies of individual fine particles extracted from a smoke have been extensively studied by X-rays and by high resolution transmission and scanning electron microscopy. In general, each particle is found to be either a single crystal or a twinned crystal, and provided it is prepared in a pure gas atmosphere it usually exhibits a well-defined crystal habit. The crystal structure, with a few exceptions, is identical to that of the bulk pure metal, and the habits, with the exception of rods or plates found in the outer zone of a candle-like smoke, are similar to those expected from the Wulff construction. Thus hexagonal metals have habits which are plates or isometric shapes of

hexagonal symmetry, f.c.c. crystals have habits bounded solely by  $\{111\}$  planes (octahedra) or  $\{111\}$  and  $\{100\}$  planes (truncated octahedra... cuboctahedra) and b.c.c. crystals have habits bounded by  $\{110\}$  planes (rhombic dodecahedra) or  $\{110\}$  and  $\{100\}$  planes (truncated rhombic dodecahedra... cubes). Multiply twinned f.c.c. particles, already mentioned, must have appreciable internal stresses because of the misfit between five (or 20) tetrahedra and a pentagonal decahedron (or an icosahedron), but below a critical size estimated at 10 nm an icosahedral particle has lower energy than the octahedron. Sufficiently small clusters are effectively large molecules rather than twinned crystallites; it is well known that a cluster of 13 spherical atoms defines an octahedron in a close-packed arrangement which cannot be achieved in a crystal, and the smallest number of spherical atoms which defines a pentagonal decahedron is only seven.

If the ultrafine particles or clusters produced by one of the above techniques can be consolidated into a continuous body, the resultant nanocrystalline material may have properties appreciably different from those of polycrystalline material of conventional grain size. Gleiter considers three methods for the deposition of the clusters on to a substrate. In the first method, the clusters are entrained in a gas jet and are thus "sprayed" at high speed on to the substrate. The fine particles stick to the substrate and to one another under conditions where no deposit is formed from larger particles; co-deposition of different particles is also possible. A second, rather specialized, method for production of nanocrystalline thin films uses ionized cluster beams, generated by the adiabatic expansion of a vapour beam through a nozzle to obtain the clusters, some of which are then ionized by an electron beam. The ionized clusters are accelerated on to the substrate by an applied field. The third and most general method is to collect the clusters, and then to consolidate them by powder metallurgical techniques. The clusters are transported by convection to a rotating, liquid nitrogen-cooled, "cold finger" in the gas condensation chamber. They accumulate as a fine powder which is periodically scraped off the finger, directed into a low pressure compaction cylinder where the powder is consolidated and then further compacted in a high pressure unit. Both compactations are carried out under ultrahigh vacuum conditions.

In some circumstances, vapour deposition on to a substrate which is maintained at a very low temperature may lead to the formation of non-crystalline rather than microcrystalline or nanocrystalline solids. Amorphous metals, also known as "metallic glasses", may be formed by the deposition of nominally pure noble and transition metals (Bennett and Wright, 1972), but the conditions are rather stringent. Alloys are easier to prepare in the glassy or amorphous state than are pure metals; this seems to apply especially to alloys of compositions around  $M_4X$ , where M is a noble or transition metal and X is either a single metalloid or a mixture of metalloids. An amorphous or glassy solid has no long-range order, but local atomic arrangements may be determined by preferred translations (i.e. near neighbours tend to be separated by a characteristic interatomic distance) and preferred symmetry (e.g. the relative dispositions of near neighbours may correspond approximately to local rotation axes). As discussed for liquids on pp. 162–166, averaged structural information about the interatomic distance may be obtained from the position of the first (diffuse) diffraction peak, and about the coordination number from the

experimental radial distribution function. Metallic glasses are believed to have liquid-like structures with a tendency, in the case of pure metals at low temperatures, to random close packing (RCP) of equi-sized, hard spheres (see pp. 162–164). The structural similarity may help to explain the formation of amorphous solid regions from severely undercooled liquids, but such regions may also form from vapours or beams. Amorphous films or multilayers may presumably be grown from the vapour because the lack of surface mobility over the substrate at the deposition temperature prevents ad-atoms from rearranging into lower energy crystalline forms. One factor favouring glass formation in the case of alloys may be a critical composition range combined with a radius ratio of the atomic species which favours clusters of fivefold symmetry. This implies that amorphous solids may be more readily obtained from alloys with compositions at or near those corresponding to Frank–Kasper phases with appropriate coordination number. Another suggested link with Frank–Kasper phases is the recently discovered formation of “quasi-crystalline” structures in certain alloys.

Until about 1984, it was generally believed that all solids are either crystalline (characterized by long-range translational symmetry) or amorphous (without any form of long-range order). However, a pioneering paper by Shechtman *et al.* (1984) first established the existence, in an alloy of composition  $\text{Al}_{86}\text{Mn}_{14}$ , of a third type of solid, now known as “quasi-crystalline”. The “quasi-crystalline” state has a structure in which rotational symmetry (exemplified, in particular, by fivefold axes) is preserved over large distances but there is no long-range translational symmetry. The Al–Mn alloy, for example, was found to have fivefold, threefold and twofold rotational axes and exhibited the full rotational symmetry of the icosahedral point group  $m\bar{3}\bar{5}$ . As discussed in Chapter 6, fivefold rotational symmetry is not compatible with any crystallographic space group. The structure thus has no translational symmetry and cannot be periodic, but the quasi-crystalline solids are nevertheless highly ordered structures.

The challenge to conventional crystallography presented by the initial discovery of the *i*-phase (*i* = icosahedral) caused much scientific interest and there are now many thousands of published papers on the structure and properties of quasi-crystals. The *i*-phase has been obtained in very many aluminium- and titanium-based alloys, and is the commonest type of quasi-crystal, but there are others; for example, a so-called decagonal phase has been identified in several aluminium-based (but not apparently in titanium-based) systems. This “*d*-phase” was discovered by Sastry *et al.* (1978) in a slowly cooled alloy of composition  $\text{Al}_{60}\text{Mn}_{11}\text{Ni}_4$ , and was later shown by Chattopadhyay *et al.* (1985) and by Bendersky (1985) to be a two-dimensional quasi-crystal with one periodic axis normal to two non-periodic axes.

Although experimental results were not available before 1984, theoretical studies relevant to the concepts of quasi-crystallinity had been published previously, especially by Penrose (1974) and by Mackay (1981, 1982). Penrose first demonstrated that two-dimensional space may be filled by a non-periodic arrangement of two differently shaped tiles, and that this tiling pattern possesses long-range orientational order with fivefold symmetry. In a two-dimensional crystal, in contrast, space may be filled by a single tile with the parallelogram shape of some unit cell; the tiles then all have the same orientation

and their aligned edges define the two-dimensional lattice. Penrose tiles have to be fitted together in accordance with certain matching rules which determine whether the resulting structure will be crystalline, random or quasi-crystalline. Although Penrose originally used kite- and dart-shaped tiles, the tiling is most easily represented by using two tiles which are both rhombus-shaped with equal sides but different interior angles, namely  $72^\circ + 108^\circ$  and  $144^\circ + 36^\circ$ . Three-dimensional tiling was first shown to be feasible by Mackay (1981) and by Ammann (see Grunbaum and Shepard, 1986). In a three-dimensional quasi-crystal, constructed from two different rhombohedral "tiles", the matching rules are more complex and have to satisfy both local and long-range conditions; this has led to suggestions that a three-dimensional quasi-crystal can seldom be perfect.

The notion of quasi-periodicity and a quasi-lattice is most readily recognized first in a one-dimensional chain where links of two different lengths  $L$  and  $S$  alternate in a Fibonacci sequence in which a recursive relationship  $L \rightarrow L + S$ ,  $S \rightarrow L$  is used to build up successive distributions of increasing periods...  $LS\dots$ , ...  $LSL\dots$ , ...  $LSLLS\dots$ , ...  $LSLLSLSL\dots$ , etc. As the period tends to infinity, the ratio of the number of  $L$  links to  $S$  links tends to the "golden mean",  $\tau = 1/2[1 + 5^{1/2}]$ . Mackay noted that the vertices of the Penrose tiles could be defined by a "quasi-lattice" with axes along five directions at successive angles of  $72^\circ$ , the spacing of vertices along each direction consisting of a non-periodic alternation of two lengths in the manner of a Fibonacci sequence. Similarly, three-dimensional structures may be described as an atomic decoration of a three-dimensional quasi-lattice.

Shectman *et al.* reported that electron diffraction patterns consisted of sharp spots with apparent tenfold symmetry, and equivalent patterns were produced by Mackay (1982) using visible light and a screen with holes in the positions of the vertices of the Penrose tiles.

Since the initial discovery of a quasi-crystalline phase, very many investigations into the structure and properties of such phases have been published and the subject is still rapidly developing. The initial discovery of the *i*-phase was very quickly followed by reports of various other phases with fivefold or related symmetries. The first to be reported, as mentioned above, was a two-dimensional *d*-phase with a periodic structure along one axis only, and planes normal to this axis exhibiting tenfold rotational symmetry. Other phases include an octagonal quasi-crystal (Wang *et al.*, 1987), a dodecagonal quasi-crystal (Ishimasa *et al.*, 1985), a pentagonal quasi-crystal (Menguy *et al.*, 1993a), a cubic quasi-crystal (Donnadieu, 1994) and a hexagonal quasi-crystal (Selke *et al.*, 1994).

One interesting way of viewing a quasi-crystalline structure is to consider a periodic structure in a higher dimensional space and then to project this structure on to the real space. Thus a two-dimensional square grid, if projected on to an irrationally oriented line, will yield the Fibonacci sequence, and a hypercube in six dimensions can yield the *i*-phase if projected on an irrationally orientated three-dimensional space. The quasi-crystalline phases are also related to Frank-Kaspar crystalline phases which contain similar atomic groupings.

The question "where are the atoms?" has caused great difficulties in theoretical descriptions of quasi-crystals. The lack of periodicity precludes the calculation of a

structure factor by normal methods, and Bak (1986) pointed out that the structure can only be tackled by calculating structure factors in six dimensions and then projecting this on to three dimensions. Cahn *et al.* (1988) have proposed a six-dimensional structural model, which can be related to three-dimensional structure in this way.

Even though the detailed atomic structure remains obscure, it is known that quasi-crystalline phases can both order and twin. Superlattice formation in the *i*-phase was demonstrated by Mukhopadhyay *et al.* (1987, 1989) and twinning via a mirror plane normal to one of the fivefold axes was studied by Ranganathan *et al.* (1989). Both ordering and twinning were reviewed by Singh and Ranganathan (1997).

## REFERENCES

- AMELINCKX, S., BONTINCK, W., DEKEYSER, W., and SEITZ, F. (1957) *Philos. Mag.* **2**, 355.  
 BAK, P. (1986) *Phys. Rev. Lett.* **56**, 861.  
 BASSETT, G. A., MENTER, J. W., and PASHLEY, D. W. (1959) *Structure and Properties of Thin Films*, eds. C. A. Neugebauer, J. B. Newkirk and D. A. Vermilyea, Wiley, New York.  
 BASSETT, G. A. (1958) *Philos. Mag.* **3**, 1042.  
 BAUSER, E. and STRUNK, H. (1981) *J. Cryst. Growth* **51**, 362.  
 BENDERSKY, L. (1985) *Phys. Rev. Lett.* **55**, 1461.  
 BENNEMA, P. and VAN LEEUWEN, C. (1975) *J. Cryst. Growth* **31**, 3.  
 BENNETT, M. R. and WRIGHT, J. G. (1972) *Phys. Status Solidi A* **13**, 135.  
 BETHGE, H. (1962) *Phys. Status Solidi* **2**, 775; (1967) *Crystal Growth*, ed. H. S. Peiser, p. 623, Pergamon, Oxford.  
 BETHGE, H., KELLER, K. W., and ZEIGLER, E. (1968) *J. Cryst. Growth* **3-4**, 184.  
 BINDER, K. and KALOS, M. H. (1980) *J. Stat. Phys.* **22**, 363.  
 BLOEM, J. and CLASSEN, W. A. P. (1983) *Philips Tech. Rev.* **41**, 60.  
 BONNISSANT, A. and MUTAFTSCHIEV, B. (1974) *J. Cryst. Growth* **24-25**, 503.  
 BURTON, W. K. and CABRERA, N. (1949) *Discuss. Faraday Soc.* **5**, 33, 40.  
 BURTON, W. K., CABRERA, N., and FRANK, F. C. (1949) *Nature (London)* **163**, 398; (1950-1) *Phil. Trans. R. Soc. London, Ser. A* **243**, 299.  
 CABRERA, N. and COLEMAN, R. V. (1963) *The Art and Science of Growing Crystals*, ed. J. J. Gilman, Wiley, New York.  
 CABRERA, N. and LEVINE, M. M. (1956) *Philos. Mag.* **1**, 450.  
 CABRERA, N. and VERMILYEA, D. A. (1958) *Growth and Perfection of Crystals*, eds. R. H. Doremus, B. W. Roberts and D. Turnbull, p. 393, Wiley, New York, Chapman and Hall, London.  
 CAHN, J. W., GRATIAS, D., and MOZER, B. (1988) *J. Phys. (Paris)* **49**, 1225.  
 CAHN, J. W. and HILLIARD, J. E. (1957) *J. Chem. Phys.* **32**, 688.  
 CHAPON, C. and HENRY, C. R. (1981) *Surf. Sci.* **106**, 152.  
 CHAPON, C., HENRY, C., and MUTAFTSCHIEV, B. (1976) *J. Cryst. Growth* **33**, 291.  
 CHATTOPADHYAY, K., RANGANATHAN, S., SJUBBANNA, G. N., and THIANGARAJ, N. (1985) *Scr. Metall.* **19**, 767.  
 COLEMAN, R. V. and SEARS, G. W. (1957) *Acta Metall.* **5**, 31.  
 DE HAAN, S. W. H., MEEUSSEN, V. J. A., VELTMAN, B. P., BENNEMA, P., VAN LEEUWEN, C., and GILMER, G. H. (1974) *J. Cryst. Growth* **24-25**, 491.  
 DONNADIEU, P. (1994) *J. Phys. I France* **4**, 791.  
 EGGINGTON, A., KING, C. S., STAUFFER, D., and WALKER, G. H. (1971) *Chem. Phys. Lett.* **26**, 820.  
 ELBAUM, C. and CHALMERS, B. (1955) *Can. J. Phys.* **33**, 196.  
 FISHER, X. (1967) *Physics* **3**, 255.  
 FORTY, A. J. (1952) *Philos. Mag.* **43**, 72.  
 FORTY, A. J. and FRANK, F. C. (1953) *Proc. R. Soc. London, Ser. A* **217**, 262.

- FRANK, F. C. (1951) *Philos. Mag.* **42**, 809; (1952) *Adv. Phys.* **1**, 91; (1953) *Philos. Mag.* **44**, 8, 856; (1958) *Growth and Perfection of Crystals*, p. 411, Wiley, New York.
- FRANK, F. C. and IVES, M. B. (1960) *J. Appl. Phys.* **31**, 1996.
- FRANKL, D. and VENABLES, J. A. (1970) *Adv. Phys.* **19**, 409.
- GIBBS, J. W. (1875–8) *Trans. Conn. Acad. Arts Sci.* **3**, 108, 343; (1906) *The Collected Works of J. Willard Gibbs*, Yale University Press, New Haven, CT.
- GILMER, G. H. (1976) *J. Cryst. Growth* **35**, 15; (1980) *Ibid.* **49**, 465.
- GILMER, G. H., GHEZ, R., and CABRERA, N. (1971) *J. Cryst. Growth* **8**, 79.
- GLEITER, H. (1989) *Prog. Mater. Sci.* **33**, 223.
- GOMER, R. J. (1957) *J. Chem. Phys.* **26**, 1333.
- GRANDQVIST, C. G. and BUHRMAN, R. A. (1976) *J. Appl. Phys.* **47**, 2200.
- GRUNBAUM, B. and SHEPARD, G. (1986) *Patterns and Tilings*, Freeman, New York.
- GUTMANAS, E. Y. (1990) *Prog. Mater. Sci.* **34**, 261.
- HAM, F. S. (1958) *J. Phys. Chem. Solids* **6**, 335.
- HARDY, A. H. (1956) *Prog. Met. Phys.* **6**, 45.
- HARRIS, S. (1989) *J. Cryst. Growth* **97**, 319.
- HARSDORFF, M. and KNABBE, E. A. (1979) *Surf. Sci.* **86**, 36.
- HENRY, C., CHAPON, C., and MUTAFTSCHIEV, B. (1976) *Thin Solid Films* **33**, L1.
- HILLIG, B. (1966) *Acta Metall.* **14**, 1868.
- HIRTH, J. P. (1959) *Acta Metall.* **7**, 755.
- HIRTH, J. P. and POUND, G. M. (1957a) *J. Chem. Phys.* **26**, 1216; (1957b) *Acta Metall.* **5**, 649; (1963) *Prog. Mater. Sci.* **11**, 1.
- HULLETT, L. D. and YOUNG, F. W. (1966) *J. Electrochem. Soc.* **113**, 410.
- INO, S. (1948) *J. Phys. Soc. Jpn.* **9**, 359; (1966) *Ibid.* **21**, 346.
- ISHIMASA, T., NISSEN, H.-U., and FUKANO, Y. (1985) *Phys. Rev. Lett.* **55**, 511.
- ISRAEL, A., PETROFF, J. E., AUTHIER, A., and MAUK, Z. (1972) *J. Cryst. Growth* **16**, 131.
- IVES, M. B. (1961) *J. Appl. Phys.* **32**, 1534.
- JACKSON, K. A. (1958) *Liquid Metals and Solidification*, p. 174, American Society of Metals, Metals Park, OH.
- JACKSON, K. A. and MILLER, C. E. (1977) *J. Cryst. Growth* **40**, 169.
- JACUCCI, G., PERINI, A., and MARTIN, G. (1983) *J. Phys. A: Math., Nucl. Gen.* **16**, 369.
- KATZ, J. L., MIRABELL, P., SCOPPO, C. Z., and VIRKLER, T. L. (1976) *J. Chem. Phys.* **64**, 382.
- KELLER, K. W. (1969) *Phys. Status Solidi* **36**, 557.
- KNABBE, E. A. and HARSDORFF, M. (1979) *Thin Solid Films* **57**, 271.
- KOSSEL, W. (1927) *Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl.* 135.
- KROHN, M. and BETHGE, H. (1979) *Thin Solid Films* **57**, 227.
- LAASER, W., DABRINGHAUS, H., and MEYER, H. J. (1983) *J. Cryst. Growth* **62**, 284.
- LANGER, J. S. and TURSKE, L. A. (1973) *Phys. Rev. A* **8**, 3230.
- LEAMY, H. J. and GILMER, G. H. (1974) *J. Cryst. Growth* **24–25**, 499.
- LEAMY, H. J., GILMER, J. H., and JACKSON, K. A. (1975) *Statistical Thermodynamics of Clean Surfaces*, ed. J. M. Blakely, p. 121, Academic Press, New York.
- LEAMY, H. J. and JACKSON, K. A. (1971) *J. Appl. Phys.* **42**, 2121.
- LEE, K., BARKER, J. A., and ABRAHAM, F. F. (1973) *J. Chem. Phys.* **58**, 3160.
- LEWIS, B. (1970) *Surf. Sci.* **21**, 273, 289.
- LEWIS, B. and ANDERSON, J. C. (1978) *Nucleation and Growth of Thin Films*, Academic Press, New York.
- LEWIS, B. and REES, G. J. (1974) *Philos. Mag.* **29**, 1253.
- LIGHTHILL, M. J. and WHITHAM, G. B. (1955) *Proc. R. Soc. London, Ser. A* **229**, 281, 317.
- LOTHE, J. and POUND, G. M. (1962) *J. Chem. Phys.* **36**, 2080.
- MACKAY, A. C. (1981) *Sov. Phys.—Crystallogr.* **26**, 517; (1982) *Physica A* **114**, 609.
- MADONNA, L. A., SCIALLI, C. M., CANIAR, L. N., and POUND, G. M. (1961) *Proc. Phys. Soc., London* **28**, 1218.
- MAR, R. W. and SEARCY, A. W. (1970) *J. Chem. Phys.* **53**, 3076.
- MARUYAMA, M. (1988) *J. Cryst. Growth* **89**, 415.

- MATTHEWS, J. W. (1979) *Dislocations in Solids*, ed. F. N. Nabarro, p. 461, North-Holland, Amsterdam.
- MENGUY, N., AUDIER, M., GUYOT, P., and VACHER, M. (1993a) *Philos. Mag.* **68B**, 593; (1993b) *Ibid.* **68B**, 597.
- MEYER, J. and STEIN, B. J. (1980) *J. Cryst Growth* **49**, 707; (1982) *Ibid.* **60**, 123.
- MUKHOPADHYAY, N. K., RANGANATHAN, S., and CHATTOPADHYAY, K. (1987) *Philos. Mag. Lett.* **56**, 121; (1989) *Ibid.* **60**, 207.
- MULLINS, W. W. and HIRTH, J. P. (1963) *J. Phys. Chem. Solids* **24**, 1391.
- NABARRO, F. R. N. and JACKSON, P. J. (1958) *Growth and Perfection of Crystals*, eds. R. H. Doremus, B. W. Roberts and D. Turnbull, p. 11, Wiley, New York, Chapman and Hall, London.
- NAKATANI, I., FURUBAYASHI, T., TAKAHASHI, T., and HANAOKA, H. J. (1987) *J. Magn. Magn. Mater.* **65**, 261.
- NIELSEN, A. (1964) *Kinetics of Precipitation*, p. 46, Pergamon Press, Oxford.
- OBREtenOV, W., BOSTANOV, V., BUDEVSKI, E., BARRADAS, R. G., and VAN DER NOOT, T. J. (1986) *Electrochim. Acta* **31**, 753; (1989) *J. Cryst. Growth* **96**, 843.
- OHARA, M. and REID, R. C. (1973) *Modelling Crystal Growth Rates from Solution*, Prentice-Hall, Englewood Cliffs, NJ.
- PASHLEY, D. W. (1956) *Adv. Phys.* **5**, 173.
- PENROSE, R. (1974) *Inst. Math. Appl. Bull.* **10**, 266.
- PERINI, A., JACUCCI, G., and MARTIN, G. (1984a) *Phys. Rev. B* **29**, 2689; (1984b) *Surf. Sci.* **144**, 53.
- PIPER, W. W. and POLICH, S. J. (1961) *J. Appl. Phys.* **32**, 1278.
- PONTIKIS, V. and SINDZINGRE, P. (1987) *Phys. Scr.* **T19**, 375.
- POUND, G. M. (1985) *Metall. Trans.* **16A**, 487.
- POUND, G. M., SIMNAD, N. T., and YANG, L. (1954) *J. Chem. Phys.* **22**, 1215.
- RANGANATHAN, S., PRASAD, R., and MUKHOPADHYAY, H. A. (1989) *Philos. Mag. Lett.* **51**, 157.
- READ, W. T., Jr. (1953) *Dislocations in Crystals*, McGraw-Hill, New York.
- REICHEL, K., LAMPERT, B., and SIEGERS, H. P. (1980) *Surf. Sci.* **93**, 159.
- REISS, H. (1977) *Nucleation Phenomena: Advances in Colloid and Interface Science*, ed. A. C. Zettlemeyer, p. 205, Elsevier, New York.
- ROBERTSON, D. and POUND, M. (1973) *J. Cryst. Growth* **19**, 269.
- ROBINSON, V. N. E. and ROBINS, J. L. (1970) *Thin Solid Films* **5**, 313; (1974) *Ibid.* **20**, 155.
- SASTRY, G. V. S., SURYANARAYANA, C., VAN SANDE, M., and VAN TENDELOO, G. (1978) *Mater. Res. Bull.* **13**, 1065.
- SCHEY, J. (1986) *Encyclopedia of Materials Science and Engineering* **6**, p. 4806, Pergamon Press, Oxford.
- SEARS, G. W. (1955) *Acta Metall.* **3**, 361, 367.
- SEARS, G. W. and COLEMAN, R. V. (1956) *J. Chem. Phys.* **23**, 635.
- SELKE, H., VOGG, U., and RYDER, R. C. (1994) *Phys. Status Solidi A* **144**, 31.
- SHARAF, M. A. and DOBBINS, R. A. (1982) *J. Chem. Phys.* **77**, 1517.
- SHECHTMAN, D., BLECH, I., GRATIAS, D., and CAHN, J. W. (1984) *Phys. Rev. Lett.* **53**, 1951.
- SINGH, A. and RANGANATHAN, S. (1997) *Prog. Mater. Sci.* **42**, 397.
- SPILLER, G. D. T., AKHTER, P., and VENABLES, J. A. (1983) *Surf. Sci.* **131**, 517.
- STEIN, B. J. and MEYER, H. J. (1980) *J. Cryst. Growth* **49**, 696.
- STOWELL, M. J. (1972) *Philos. Mag.* **26**, 349, 361; (1974) *J. Cryst. Growth* **24-25**, 45.
- STOWELL, M. J. and HUTCHINSON, T. E. (1971) *Thin Solid Films* **8**, 41.
- STRANSKI, I. N. (1928) *Z. Phys. Chem.* **136**, 259.
- SUNAGAWA, I. and BENNEMA, P. (1981) *J. Cryst. Growth* **53**, 490.
- SUREK, T. (1972) *J. Cryst. Growth* **13-14**, 19.
- SUREK, T., CHALMERS, B., and MLAVSKY, A. I. (1977) *J. Cryst. Growth* **42**, 453.
- SUTTON, A. P. (1986) *Encyclopedia of Materials Science and Engineering* **2**, p. 782, Pergamon Press, Oxford.
- SWENDSEN, R. H., KORTMAN, P. J., LANDAU, D. P., and MILLER-KRUMBHAR, H. (1976) *J. Cryst. Growth* **35**, 73.
- TROFIMOV, I. (1981) *J. Cryst. Growth* **54**, 211.
- TSANG, W. T. (1987) *J. Cryst. Growth* **81**, 261; (1989) *Ibid.* **95**, 121.

- USHER, B. F. and ROBINS, J. L. (1982) *Thin Solid Films* **90**, 15.
- UYEDA, R. (1991) *Prog. Mater. Sci.* **35**, 1.
- VAN BELJEREN, H. (1977) *Phys. Rev. Lett.* **38**, 993.
- VAN DER HOEK, B., VAN DER EERDEN, J. P., and BENNEMA, P. (1982a) *J. Cryst. Growth* **56**, 621; (1982b) *Ibid.* **58**, 365.
- VAN DER MERWE, J. H. (1973) *Treatise on Materials Science and Technology*, Academic Press, New York.
- VELFE, H. D. and KROHN, M. (1982) *Thin Solid Films* **98**, 125.
- VELFE, H. D., STENZEL, H., and KROHN, M. (1982) *Thin Solid Films* **98**, 115.
- VENABLES, J. A. (1973) *Philos. Mag.* **27**, 697.
- VENABLES, J. A., SPILLER, J. A. T., and HANBUCKEN, M. (1984) *Rep. Prog. Phys.* **43**, 399.
- VERMA, A. R. (1953) *Crystal Growth and Dislocations*, Butterworths, London.
- VINCENT, R. (1971) *Proc. R. Soc. London, Ser. A* **321**, 53.
- VIOLA, M., VAN WORMER, K., and BOTSARIS, G. (1979) *J. Cryst. Growth* **47**, 127.
- VOLMER, M. and FLOOD, H. (1934) *Z. Phys. Chem., Abt. A* **170**, 273.
- WALTON, D. (1962) *J. Chem. Phys.* **27**, 218.
- WALTON, D. and RHODIN, T. N. (1963) *Metal Surfaces*, p. 259, American Society for Metals, Metals Park, OH.
- WANG, N., CHEN, H., and KUO, K. (1987) *J. Cryst. Growth* **69**, 1010.
- WEBB, W. W. (1958) *Acta Metall.* **7**, 4.
- WEEKS, J. D. and GILMER, G. H. (1976) *J. Cryst. Growth* **33**, 21; (1979) *Adv. Chem. Phys.* **40**, 157.
- WINTERBOTTOM, W. L. (1967) *Acta Metall.* **15**, 303.
- WOODRUFF, D. P. (1986) *Encyclopedia of Materials Science and Engineering* 6, p. 4803, Pergamon Press, Oxford.
- YAGI, K., TAKAYANAGI, K., KOBAYASHI, K., OSAKABE, N., TANISHIRO, Y., and HONJO, G. (1979) *Surf. Sci.* **86**, 174.
- YATSUYA, S., TSUKASAKI, Y., MIHAMA, K., and UYEDA, R. (1978) *J. Cryst. Growth* **45**, 490.
- YATSUYA, S., YANAGIDA, A., YAMAUCHI, K., and MIHAMA, K. (1984) *J. Cryst. Growth* **70**, 536.
- ZINNSMEISTER, G. (1968) *Thin Solid Films* **2**, 497; (1969) *Ibid.* **4**, 363.

## CHAPTER 14

# *Solidification and Melting*

### 64. THE SOLIDIFICATION OF PURE METALS

It is usually assumed that, after a brief initial transient, a pure liquid metal at a temperature below its equilibrium melting/freezing point will form stable crystalline nuclei at a rate given by eqn. (49.11), and these nuclei will subsequently grow into larger crystals. The overall transformation rate is thus, in principle, dependent on two separate processes, as is also the rate of crystal growth from the vapour. Once a stable nucleus has formed in a supercooled liquid, however, its growth rate is usually very large and may be limited only by the rate at which the released latent heat of solidification can be removed from the vicinity of the solid–liquid interface. In some circumstances, this latent heat may raise the temperature of the remaining liquid mass above that at which the first solid nuclei were formed, and further nucleation may then be inhibited. At the other extreme, if the thermal conditions permit the heat to be removed rapidly, the growth time is negligible and the time taken for solidification is determined primarily by the rate of formation of the first nuclei. Under such conditions, the applicability of the classical theory of quasi-steady-state nucleation to calculate the overall rate of solidification seems unrealistic as the time taken to establish the quasi-steady-state distribution of embryos of various sizes is appreciably greater than the time taken to form each nucleus once the steady state has been attained. The theory of transient nucleation (Section 50) thus implies that the first nucleus may take longer to form than is predicted by eqn. (49.11). This difficulty has generally been ignored and experimental results have been compared with the predictions of eqn. (49.11); this may be justified because of the very large temperature dependence of both steady-state and transient nucleation rates.

It has long been known that liquids may be cooled to temperatures well below the equilibrium freezing point before solidification begins. As first shown by Fahrenheit for water, a particular specimen begins freezing at some characteristic undercooling which is reproducible in successive experiments with that sample. This is expected from the theory of Chapter 10 because the nucleation rate changes so rapidly with temperature (see Fig. 10.3) that “appreciable” nucleation (however vaguely defined) effectively corresponds to a very small temperature range. However, experiments with different, and apparently indistinguishable, samples of the same liquid often result in widely different values for the amount of supercooling required to freeze the sample. When the liquid volume is

reasonably large (say about  $10^{-6} \text{ m}^3$  or more), freezing most frequently begins at a temperature only slightly below the equilibrium temperature, but occasional samples may supercool by  $100^\circ\text{C}$  or more. The sample differences are attributed to very small variations in their impurity contents, as catalysed nucleation on impurity particles is very sensitive to the value of the contact angle  $\theta$  [eqn. (51.7)]. Many experiments have established that the freezing of ordinary bulk liquids commonly begins either on suspended solid impurities or on the walls of the solid container.

The near impossibility of controlling minute amounts of accidental impurities is the main difficulty in experimental work performed to test the validity of the theory of homogeneous nucleation. In principle, there are two methods by which the impurities might be removed and homogeneous nucleation ensured. The first method is simply to filter the liquid in an attempt to remove all solid particles. This may be possible if the growth rate is not too large as copious nucleation will then be required to effect solidification. For metals and other liquids of high fluidity, however, each single nucleus may grow rapidly and, in extreme cases, the whole liquid may solidify as a single crystal when only one nucleus has formed. It is clearly almost impossible to filter so efficiently that not a single potential nucleating agent remains, and the practical difficulties in filtering liquid metals are considerable. Attention has therefore been concentrated on the second technique, which is to divide the liquid into a number of small droplets.

The small droplet technique utilizes the obvious result that if the liquid can be divided into many more droplets than the number of catalytic impurities which it originally contained, most of the droplets must be free of such impurities. Early work by Mendenhall and Ingersoll (1908) showed that small droplets of gold, platinum or rhodium supercooled much more than the normal amount, but the significance of the result was not then appreciated. The first application of the small droplet technique to the kinetics of solidification was made by Vonnegut (1948), who worked with tin. The method was subsequently applied to mercury and other metals by Turnbull (1950a) and by other workers.

The freezing of small droplets may be studied by careful observation of individual droplets or by measuring various properties of an aggregate of such droplets. The aggregate method, introduced by Vonnegut and extended by Turnbull, gives a much more detailed and quantitative comparison of theory with experiment, but the analysis is more complex. Vonnegut used a foaming agent to emulsify the liquid, so that the individual droplets in the aggregate were separated from each other by an inert film. It was then possible to measure the rate of solidification at various temperatures by monitoring volume changes with a dilatometer. If the majority of droplets do not contain accidental impurity particles, their solidification must require either homogeneous nucleation or heterogeneous nucleation catalysed by the substance which is coating the droplets. Vonnegut's droplets were  $1\text{--}10 \mu\text{m}$  in diameter, and the assembly contained about  $10^{11}$  particles separated by oxide films. Measurements were possible only for a very restricted range of supercooling,  $\Delta T^- = 100\text{--}115^\circ\text{C}$ . Turnbull's first measurements were made with mercury droplets coated with mercury stearate, and measurable transformation rates were found over the temperature range  $\Delta T^- = 59\text{--}63^\circ\text{C}$ .

If eqn. (49.14) is valid, a plot of  $\log {}^vI$  against  $1/T(\Delta T)^{-2}$  should be a straight line, the slope of which gives a value of  $\sigma$ , and the intercept on the  $\log {}^vI$  axis gives the pre-exponential constant. This is a reasonable experimental test of the theory if the variation of nucleation rate with temperature can be deduced from the isothermal transformation curves.

Consider an assembly of droplets, each of volume  $v$  and each having the same quasi-steady-state nucleation rate per unit volume,  ${}^vI$ , and a growth rate so rapid that each droplet effectively solidifies as soon as it acquires a stable solid nucleus. In a small time interval  $dt$ , a fraction  ${}^vIvdt$  of the droplets which are still liquid will solidify. Hence the volume fraction,  $\zeta$ , which has solidified at time  $t$  after the beginning of the phase change is given by

$$d\zeta/dt = {}^vIv(1 - \zeta) \quad (64.1)$$

so that if  $v$  is known,  ${}^vI$  can be evaluated from the slope of the experimental  $(1 - \zeta)$  curve at any fixed value of  $\zeta$ . Turnbull assumed that this equation may be applied to a distribution of droplet sizes if  $v$  is taken to be the root-mean-square droplet volume. Although, in fact, the experimental  $(1 - \zeta)$  curves were not of the simple form predicted by eqn. (64.1), the correct temperature dependence of  ${}^vI$  should still be obtained by evaluating the apparent  ${}^vI$  at different temperatures at a constant fractional amount of solidification (see below), provided  ${}^vI$  is the same for all drops. Turnbull's analysis of his own and Vonnegut's results in this way gave values of  $10^{36}$ – $10^{37} \text{ m}^{-3} \text{ s}^{-1}$  for the pre-exponential term in the nucleation rate, in very good agreement with the theoretical estimate of  $10^{39} \text{ m}^{-3} \text{ s}^{-1}$  of eqn. (49.12). This apparent confirmation of the theory is, however, unsatisfactory, and later work by Pound and La Mer (1952) showed that the results for tin are actually inconsistent with the assumption of a single nucleation frequency. In a more thorough analysis, attention has to be paid to the size distribution, as Vonnegut recognized. This was undertaken by Pound and La Mer for tin and by Turnbull (1952a) for mercury.

In an assembly containing  $N_r$  droplets of radius  $r$  and volume  $v_r$ , the total volume of such droplets is  $V_r = N_r v_r$ , where  $N_r$  and  $V_r$  may be either continuous or discontinuous functions of  $r$ . Equation (64.1) will now give the rate of solidification for droplets of any one size which may be written

$$d\zeta_r/dt = I_r(1 - \zeta_r)$$

giving

$$\zeta_r = 1 - \exp(-I_r t) \quad (64.2)$$

where  $I_r$  is the total nucleation rate in a droplet of size  $r$ . The volume fraction of the whole assembly which has solidified after time  $t$  is  $\zeta = \sum V_r \zeta_r / V$ , so that

$$\zeta = 1 - \sum (V_r/V) \exp(-I_r t)$$

or

$$\zeta = 1 - \int (V_r/V) \exp(-I_r t) dr \quad (64.3)$$

the sum or integral corresponding to discontinuous or continuous distributions of droplet sizes respectively.

In an experiment, the distribution function  $V_r/V$  and the isothermal  $(1 - \zeta)$  curves may be measured. If only  $I_r$  in eqn. (64.3) is a strong function of temperature, the isothermal transformation curves should differ from each other only by constant factors in the exponential term, and they may all be reduced to a single isothermal by finding these factors. The curve may then be compared with theoretical curves obtained from eqn. (64.3) and the measured  $V_r$  on the assumptions either of random volume nucleation or catalysed nucleation on the droplet surface of area  $o_r$ . The two possibilities give

$$\begin{aligned} I_r &= {}^v I_r \\ I_r &= {}^o I_r \end{aligned} \quad (64.4)$$

From eqns. (64.3) and (64.4), it follows that  $d\zeta/dt$  is proportional to  ${}^v I$  or  ${}^o I$  respectively. This is the justification for Turnbull's assumption above that the temperature dependence of the nucleation rate obtained with eqn. (62.1) is substantially correct although the exact values of  $I$  are uncertain if the size distribution is not measured. The result is valid only if there is a single characteristic nucleation frequency  $I_l$  for each droplet size and the later work showed the necessity of establishing this experimentally in any particular instance.

Comparison of the theoretical curves obtained from eqns. (64.3) and (64.4) with the corresponding experimental isotherms indicates (though not with complete certainty) whether volume or surface nucleation is occurring. From the experimental temperature dependence of  $I$ , the numerical constant in the appropriate equation for  ${}^v I$  or  ${}^o I$  can then be found, and this should confirm the provisional conclusion.

In his experiments on mercury, Turnbull found that the range of temperatures over which solidification took place at a measurable rate varied with the nature of the surface film. Greatest supercooling (78 C) was found with droplets coated with mercury laureate or mercury benzoate, and the initial interpretation was that nucleation in these cases is homogeneous. A careful analysis of the results for the mercury laureate coating showed that  $I_r$  was proportional to  $v_r$  and not to  $o_r$ , and the slope of the best linear plot of  $\ln I$  versus  $1/T(\Delta T^-)^2$  gave a value of  $31 \text{ mJ m}^{-2}$  for the surface free energy. The critical nucleus at  $-118^\circ \text{C}$  would then contain about 830 atoms and the only inconsistency with the theory arose from the value of the pre-exponential factor of the nucleation equation, which was found to be about  $10^{48} \text{ m}^{-3} \text{ s}^{-1}$ . By specifically considering the other terms in eqn. (49.9), Turnbull obtained a theoretical value of  $10^{41} \text{ m}^{-3} \text{ s}^{-1}$ , which is slightly higher than the estimate of eqn. (49.12) but is still much smaller than that indicated by the experimental results. This discrepancy cannot be due to the assumption of steady-state rather than transient nucleation as the latter would predict a lower initial nucleation rate. One possible reason is that the nucleus has a structure differing slightly from the equilibrium structure of solid mercury; another is that the surface free energy and/or the entropy of melting may be temperature-dependent. The experimental and theoretical values would agree if the nucleus were a metastable phase with a melting point about  $12^\circ \text{C}$  lower than that of bulk mercury, but this is now considered to be improbable.

Quantitative data which seem to indicate homogeneous nucleation from the liquid have also been obtained for gallium (Miyazawa and Pound, 1974). Droplets dispersed in two different ways gave results with differences only slightly greater than those expected from the estimated errors, but the mean value of the pre-exponential factor ( $10^{46} \text{ m}^{-3} \text{ s}^{-1}$ ) is again larger than the theoretical prediction, whilst the apparent surface free energy of  $\sim 68 \text{ mJ m}^{-2}$  is of the expected order of magnitude. Miyazawa and Pound pointed out that this discrepancy has been found only for metals and for water; there are reasonably good agreements between experimental and theoretical pre-exponential factors for hydrocarbons (Turnbull and Cormia, 1961) and, for some polymeric materials, the calculated value is larger than that measured experimentally.

As mentioned on p. 424, Lothe and Pound (1962) suggested that discrepancies as large as  $10^6$ – $10^7$  in the predicted and experimental rates might be due to the “non-classical” factor [eqn. (47.6)], i.e. in this example to the hindered rotation and translation of the crystallite nucleus in the supercooled liquid. Miyazawa and Pound, however, accepted that this is very improbable, and pointed out that it would imply that only mercury, gallium and water among the materials investigated possess appreciable Brownian motion. They considered instead the possibility that  $\sigma^{\text{sl}}$  and  $\Delta h^{\text{sp}}$  both vary with temperature in a manner which preserves the approximate linear relationship between  $\ln I$  and  $1/T(\Delta T^-)^2$ . In particular, the experimental results for mercury and gallium can be reconciled with the theory of homogeneous nucleation if  $\sigma^{\text{sl}}$  increases linearly with temperature over the small range in which kinetic measurements may be made. The assumed temperature dependence corresponds to a (negative) interfacial entropy of about  $-0.055$  and  $-0.09 \text{ mJ m}^{-2} \text{ K}^{-1}$  for gallium and mercury respectively, and both the sign and magnitude are consistent with theoretical estimates for metals (see, e.g., Ewing, 1971). There is, however, no other experimental evidence to support the Miyazawa Pound suggestion, and the improved agreement between experimental and theoretical values of the pre-exponential term is obtained at the expense of rather poorer agreement in the values of this term for heterogeneous nucleation (see below).

Quantitative observations on aggregates of small drops can readily be applied only to materials of low melting point. A second method of investigation, used by Turnbull and Cech (1950), is the microscopical observation of the solidification of single droplets. The theoretical equations for nucleation and the experimental results on aggregates both indicate that a droplet will solidify suddenly in a small, well-defined temperature range, and droplets of similar size which contain the same catalytic nucleating agents will all solidify at about the same temperature. Microscopical observation of solidification may be used for metals of both high and low melting points and was applied by Turnbull and Cech to a large number of metals using droplets of 10–100  $\mu\text{m}$  diameter. For each metal, the collected results from many individual observations gave a spectrum of discrete values of  $\Delta T^-$  at which solidification occurred. In most cases, a large proportion of the droplets solidified at the lowest temperature, and this was originally believed to be the temperature at which the supercooling is sufficiently large to give an appreciable rate of homogeneous nucleation. The smaller amounts of supercooling at which other droplets solidified were similarly interpreted as the temperatures at which various nucleation catalysts become effective.

Experiments on single droplets may be compared with techniques which give similar information on emulsified aggregates (Rasmussen and Loper, 1975) or on droplet suspensions produced by sudden expansion of vapour in a cloud chamber (Cwilong, 1945; Schaeffer, 1948; Thomas and Staveley, 1952). In each case, only one experimental parameter is measured, namely the temperature at which nucleation becomes rapid under given conditions. The predictions of the theory thus cannot be tested as fully as is possible when the actual nucleation rates are measured.

The results obtained by Turnbull and Cech indicated that for most metals the maximum supercooling observed in 50  $\mu\text{m}$  droplets is approximately proportional to the melting point on the absolute scale. For metals with close-packed crystal structures (except lead), the ratio  $\Delta T^-/T^{\text{ls}}$  was equal to  $0.18 \pm 0.02$ ; higher values were found for gallium and mercury and lower values for lead and antimony. The nearly constant ratio was explained using the hypotheses that the entropy of melting is nearly the same for all metals and that the liquid–solid surface free energy  $\sigma^{\text{sl}}$  is proportional to the latent heat  $\Delta h^{\text{sl}}$ . Assuming that  $\sigma^{\text{sl}}$  is isotropic, eqn. (49.14) applied to a spherical nucleus may be written

$$(\sigma^{\text{sl}})^3 = (3kT/16\pi)(\Delta h^{\text{sp}}/\nu^{\text{l}})^2(\Delta T^-/T^{\text{ls}})^2 \ln(10^{39}/{}^{\nu}I) \quad (64.5)$$

where  ${}^{\nu}I$  (in  $\text{m}^{-3}\text{s}^{-1}$ ) is the only quantity on the right-hand side of the equation which is not directly measurable. Turnbull estimated that, at the maximum undercooling observed, the time taken for a nucleus to form was between  $10^{-2}\text{s}$  and  $1\text{s}$ , giving  ${}^{\nu}I = 2 \times 10^{11} \text{m}^{-3}\text{s}^{-1}$  in a 50  $\mu\text{m}$  droplet. The uncertainty in  ${}^{\nu}I$  is very large, but the error in  $\sigma^{\text{sl}}$  is much smaller; a factor of  $10^6$  in  ${}^{\nu}I$ , for example, changes  $\sigma^{\text{sl}}$  by only about 7%.

A possible test of the nucleation theory is to compare values of  $\sigma^{\text{sl}}$  deduced from eqn. (64.5) with measurements of the surface energy of small droplets made by some independent method. Unfortunately, there are few measurements of either droplet or macroscopic surface free energies, and the assumption, made for many years, that at the maximum supercoolings the classical experiments of Turnbull and Cech involved genuine homogeneous nucleation rested mainly on the self-consistency of the results for different metals and the (misleading) fact that  $\sigma^{\text{sl}}$  deduced from eqn. (64.5) was always of the expected order of magnitude. There is now abundant evidence that this interpretation, emphasized in the first edition of this book, was erroneous, and that most metals may be supercooled by appreciably greater amounts than the maximum values of Turnbull and Cech.

Takagi (1954) obtained the first contradictory results in his studies of the freezing and melting of thin films of lead, tin and bismuth. The films were regarded as aggregates of small particles, 1–100 nm in diameter, and structural changes were monitored by electron diffraction. Lead films of 50 nm thickness gave a supercooling in agreement with the Turnbull–Cech results but, for a 5 nm film, a supercooling of  $\sim 157^\circ\text{C}$  was observed whilst that predicted from the Turnbull–Cech results was only  $\sim 109^\circ\text{C}$ . Similar discrepancies were found with 5 nm films of tin and bismuth.

As the homogeneous nucleation rate is proportional to the volume, large liquid masses are expected, in the absence of heterogeneous nuclei, to supercool less than small droplets. Turnbull's results on mercury, for example, imply that a liquid volume of  $10^{-6}\text{m}^3$  will

supercool by up to about 80% of the maximum value found in 4  $\mu\text{m}$  droplets, and there will be similar factors for other metals. Large volumes usually supercool very much less than this, indicating the presence of effective nucleation catalysts, but very large supercooling has been observed. Walker (see Metallurgical Society of AIME, 1961) was able to cool  $\sim 400$  g of nickel by about  $0.2 T^{\text{ls}}$ , and Fehling and Schell (1962) managed to supercool 2–20 g specimens of a number of metals to about the same extent as that obtained in Turnbull and Cech's droplets. Still larger supercooling in 1 g levitation-melted samples was reported by Shiraishi and Ward (1964) and by Gomsall *et al.* (1966), the value of  $\Delta T^-$  for nickel and iron reaching  $480^\circ\text{C}$  and  $420^\circ\text{C}$  respectively, compared with  $319^\circ\text{C}$  and  $295^\circ\text{C}$  reported by Turnbull and Cech. Similarly, Powell (1965) supercooled silver melts of up to 500 g by up to  $250^\circ\text{C}$  compared with the small droplet value of  $227^\circ\text{C}$ , and Rasmussen and Loper (1975) reverted to emulsion techniques and found  $\Delta T^-$  values of up to  $130^\circ\text{C}$  and  $160^\circ\text{C}$  for tin and bismuth, compared with the old single droplet maximum values of  $118^\circ\text{C}$  and  $90^\circ\text{C}$  respectively. At the time of writing, the largest supercoolings for these two metals are  $157^\circ\text{C}$  and  $227^\circ\text{C}$ .

Various supercooling maxima reported for bismuth, tin, lead, gallium and mercury by different authors are summarized in Table X, which illustrates that in most metals the maximum supercooling is now about twice the Turnbull–Cech values. There is again a remarkable constancy (possibly coincidental?) in the maximum values of the ratio  $\Delta T^-/T^{\text{ls}}$ , which falls in the range 0.37–0.42 for all the above metals with the single exception of gallium, for which Perepezko (1980) reports the remarkable value of 0.5.

Other possible tests of nucleation theory arise from the linear relation between  $\ln d$  and  $(\Delta T^-/T^{\text{ls}})^2$  for droplets of diameter  $d$  which is predicted from eqn. (64.5). This was investigated by Takahashi and Tiller (1969) using their own data on bismuth, tin and lead particles with diameters varying from 10 to 2000  $\mu\text{m}$ . The data were consistent with this theoretical relation only in the upper part of the size range, but the results for all three metals together with the Turnbull–Cech results could be fitted to a single linear plot of  $\ln d$  against either  $\Delta T^-$  or  $\Delta T^-/T^{\text{ls}}$  (the latter gives a better fit) over the whole size range. These relations are not consistent with homogeneous nucleation theory, and Takahashi and Tiller developed a theory based on the two-dimensional nucleation of successive atomic layers on surface patches of critical size.

Actually, extrapolation of the Takahashi–Tiller plot to very small particle sizes would lead to predictions of still larger supercoolings than those reported by Takagi. Stowell *et al.* (1970) examined lead deposits with particle sizes of 10–300 nm in very high vacuum conditions in an electron microscope and found that the supercooling was very dependent on the substrate. With a  $\text{Mo}_2\text{S}$  substrate, the liquid state never persisted below  $230^\circ\text{C}$  whereas, with a carbon substrate, temperatures as low as  $90^\circ\text{C}$  were attained before solidification began. These results represent respectively smaller and larger supercooling than that predicted from extrapolation of the Takahashi–Tiller curve (for which the substrate was quartz), and in addition there was no significant dependence of supercooling on particle size in the range used by Stowell *et al.*

The conclusion from all these experiments must be that homogeneous nucleation of the solid from the liquid is even more elusive than formerly believed. Whilst not all of the

TABLE X. REPORTED VALUES OF MAXIMUM LIQUID UNDERCOOLING, BEFORE THE ONSET OF SOLIDIFICATION, IN SEVERAL PURE METALS (after Moore *et al.* 1990)

Metal	Melting point (K)	Undercooling (K)	Undercooling/melting point	Reference
Bi	544	90	0.17	1
		100	0.18	2
		173	0.32	3
		227	0.42	4
Sn	505	105	0.21	5
		122	0.24	2
		175	0.35	3
		187	0.37	4
Pb	600	67	0.11	1
		80	0.13	6
		153	0.25	6
		240	0.40	2
Ga	303	76	0.25	7
		99	0.33	2
		102	0.34	8
		150	0.50	4
Hg	234	52	0.22	3
		58	0.25	4
		88	0.38	1

Key to references in Table X.

1. Turnbull and Cech (1950)

2. Scripov (1973/4)

3. Perepezko *et al.* (1970)

4. Perepezko (1980)

5. Vonnegut (1948)

6. Turnbull (1950b)

7. Stowell (1970)

Perepezko and Paik (1984)

Miyazaki and Pound (1970)

metals examined in the single droplet technique have been re-investigated, there are strong grounds for believing that the maximum supercoolings in some or all of these experiments did not correspond to homogeneous nucleation, and the early conclusion that, for homogeneous nucleation,  $\Delta T / T^{\text{ls}} = 0.2$  has been shown to be erroneous for both small and large samples. The detailed measurements of nucleation rate made by Turnbull for mercury and by Miyazawa and Pound for gallium may be interpreted with more confidence than measurements of supercooling alone and, especially for mercury, the observations at the lowest temperatures may indeed represent homogeneous nucleation. This nucleation, however, may be transient rather than quasi-steady-state.

Measurements of the lowest temperature of rapid solidification may be unreliable indicators of homogeneous nucleation, but they at least enable a lower limit for the effective surface free energy to be set via eqn. (64.5). In a very few cases, values of surface free energy deduced in this way may be compared with experimental surface free energies obtained by other techniques. One procedure involves measurements of dihedral angles at

liquid–solid triple junctions in alloys equilibrated in a two-phase liquid plus solid region, and then quenched to room temperature. From the angles (see Section 35), the ratio of the solid–liquid interfacial free energy to the energy of what in most cases will be a random, high angle grain boundary may be determined. In a survey of such measurements, Miller and Chadwick (1967) concluded that the (extrapolated) value of this ratio for many pure metals is about 0.45, and on this basis they suggested that the values of  $\sigma^{\text{sl}}$  deduced by Hollomon and Turnbull (1953) from the early supercooling data should be multiplied by a factor of about 1.5 to give correct liquid–solid interfacial energies. Absolute measurements of  $\sigma^{\text{sl}}$  become possible if the grain boundary is a symmetrical tilt boundary in a pure metal, so that its energy may be computed (see Section 37) from the grain misorientation measured in a thin foil specimen which is also used to measure the dihedral angle at the temperature of measurement (Glicksman and Vold, 1969). An alternative method (Jones and Chadwick, 1970) depends on the curvature of the solid–liquid interface in the vicinity of a grain boundary groove; interpretation in terms of a theory due to Bolling and Tiller (1961) gives an absolute value of  $\sigma^{\text{sl}}$ .

Comparisons of  $\sigma^{\text{sl}}$  deduced from nucleation data with other measurements indicate that the nucleation results represent minimum estimates as there is no certainty that homogeneous nucleation was obtained in any of the experiments and there are also many other uncertainties, such as the possible temperature dependence of  $\sigma^{\text{sl}}$  (known to be large for water) and the magnitude of its anisotropy. The agreement between the nucleation results and the absolute measurements is nevertheless remarkably good and shows, for example, that Stowell's value of  $\sigma^{\text{sl}} = 69 \text{ mJ m}^{-2}$  for lead is much more consistent with the direct measurements than is the old value of  $33 \text{ mJ m}^{-2}$  which Turnbull obtained.

In the above discussion,  $\exp(-\Delta_{\text{a,g}}/kT)$  in eqn. (49.11) has been treated as effectively constant, but it is clear that, at sufficiently low temperatures, this factor must become large and the nucleation rate will decrease again. If  $\Delta_{\text{a,g}}$  is sufficiently large it should be possible, in principle, to increase the cooling rate to a value which prevents crystallization altogether, and the structure at low temperatures is then that of an amorphous material (a glass). This is what happens at quite small cooling rates in silicate glasses, chalcogenides and some other non-metallic materials which form viscous melts at or just below the freezing temperature. In contrast, the very high atomic mobility in liquid metals makes it virtually impossible to suppress crystallization by normal cooling. However, the development, initially by Duwez (1950), of splat cooling and other very rapid solidification techniques has enabled many metallic alloys to be transformed directly from the liquid to the amorphous state and these alloys are sometimes loosely described as "glassy metals". The now very important topic of rapid solidification is discussed briefly in Chapter 24.

Experiments on the heterogeneous nucleation of the pure solid from the liquid phase have also been made either by measuring the kinetics of solidification of an aggregate of small droplets or simply the temperatures at which individual droplets solidify. Turnbull's thorough and detailed studies on aggregates of mercury droplets coated with films of mercury acetate showed that the nucleation frequency was proportional to the surface area

of a droplet rather than to its volume. These results could be described by eqn. (51.7) within experimental error, and the derived values of the contact angle [eqn. (51.3)] and the pre-exponential term were  $72^\circ$  and  $10^{29} \text{ m}^{-2} \text{ s}^{-1}$  respectively. The pre-exponential term is in good agreement with the theoretical value, but this agreement is reduced if Miyazawa and Pound's assumption that  $\sigma^{\text{sl}}$  is temperature-dependent is incorporated into the theory. Using the linear dependence which they calculated from Turnbull's results for homogeneous nucleation, they showed that the experimental value is then too small by a factor of  $10^3$  rather than  $10^{0.6}$ , and the contact angle  $\theta$  is reduced to  $60^\circ$ . A discrepancy of  $10^3$  is outside the estimated experimental error but may not be unreasonable in view of the other uncertainties in the parameters used to deduce the theoretical value.

Other experiments were not so easy to interpret. Aggregates of mercury droplets coated with mercury iodide supercooled much less than the maximum amount observed with the droplets coated with mercury stearate, but the nucleation frequency was nevertheless proportional to the volume of a droplet rather than to its area. A suggested explanation is that nucleation is catalysed by solid mercury iodide particles which are distributed throughout the volume of each droplet, rather than only on its surface. However, the derived pre-exponential factor of the surface nucleation rate was extremely high, in excess of  $10^{70}$ . A possible explanation is the nucleation of a metastable mercury structure, or even of a mercury iodide, with a lower melting point.

Detailed analysis of the results for mercury droplets coated with other compounds and of tin droplets coated with oxide (Pound and La Mer, 1952) showed that the nucleation rate was not a single function of  $\Delta T$ . This means that the efficiency of the catalyst responsible for nucleation varies considerably in different droplets, and this has been explained quantitatively by Turnbull (1952b) using a hypothesis that the surface coating forms "patches" rather than a continuous film. If the patches have a size distribution, some of them may be smaller than the critical radius of a homogeneous nucleus at a particular temperature, whilst others may be larger. When the temperature at which nucleation on the flat surface of the substrate becomes rapid is reached, each patch will acquire a flat nucleus of solid. This nucleus will be supercritical for the liquid droplet; however, this is so only if the patch radius exceeds that of a homogeneous nucleus; patches smaller than this cannot activate solidification of the droplet unless cooled to a lower temperature. A rather similar theory of surface patches was advanced by Takahashi and Tiller (1969) to explain their results on the variation of  $\Delta T^*$  with droplet size.

The behaviour of an assembly of droplets coated with patches depends on the average number of patches per droplet with a size exceeding that of the critical homogeneous nucleus at the patch nucleation temperature. If this number is appreciably greater than unity, almost all of the droplets will solidify at this temperature (more strictly in a narrow temperature interval), but if the number is less than unity, many droplets will remain liquid to lower temperatures. As the temperature is reduced, patches of smaller and smaller sizes become supercritical and can grow, so that there are apparently many nucleation frequencies in otherwise identical droplets. The effect of the smaller patches is to produce "athermal" nuclei—so that the amount of transformation below the temperature at which heterogeneous nucleation is rapid is a function of temperature but not time. The formal

analogy to athermal martensite formation is interesting. By assuming a reasonable patch size distribution, Turnbull was able to explain his anomalous results.

Controlled experiments on the heterogeneous nucleation of solid tin were conducted by Glicksman and Childs (1962). The cleaned nucleating agent was introduced into a large mass of liquid tin above its freezing point and the supercooling which initiated freezing was then measured. The contact angle could be deduced by making the usual assumption about the magnitude of the nucleation rate at the maximum supercooling observed. Glicksman and Childs found that metals were more effective catalysts for the nucleation of tin than non-metallic materials; the maximum supercooling observed with a metallic surface was about 180°C, but this was obtained with an aluminium substrate which probably had an adherent film of aluminium oxide to prevent metal-to-metal contact. An anisotropic effect was found by using single crystals of yttrium as catalysts. With only the basal plane exposed to the liquid, the measured supercooling was 8°C, but this was reduced to 6°C when the prism planes were also exposed. This result is in agreement with the hypothesis of Turnbull and Vonnegut (1952) and many other workers that the nucleation efficiency of a substrate is improved by better lattice matching. However, when different metals are compared, chemical effects seem more important; for example, platinum gave a smaller supercooling than the structurally better-matched silver. The importance of chemical effects has been confirmed in more recent experiments using a different method of studying solidification in small droplets.

In the embedded droplet technique, first used by Wang and Smith (1950), suitable thermomechanical treatment is used to produce a microstructure in which low melting point particles are embedded in a higher melting point matrix. If the embedded particles are melted and then cooled, their solidification is governed largely by heterogeneous nucleation on either impurity particles or on the high melting point matrix. If the dispersion is sufficiently fine, the majority of the particles will not contain any nucleation catalysts other than the higher melting point matrix. This method gives better reproducibility in measurements of supercooling (Southin and Chadwick, 1978; Moore *et al.*, 1990), as the environment of each droplet particle is controlled and the interfaces are not exposed to the atmosphere. Moreover, the required microstructure can often be obtained rather readily by chill casting or rapid solidification of a sample initially held at constant temperature in an equilibrium liquid plus solid two-phase field. In practice, this means utilizing the alloy systems which have eutectic or monotectic reactions. One rather obvious disadvantage of the method is that only alloy droplets in a limited composition range may be studied; in particular, the freezing of pure metals cannot be investigated in this way unless the equilibrium diagram shows that there is zero solubility (i.e. complete immiscibility).

Southin and Chadwick studied a number of such systems; after remelting the embedded droplets, the course of their resolidification on continuous cooling or on isothermal holding was followed by thermal analysis methods. In many cases, they found supercooling appreciably greater than the maximum values reported by Turnbull and Cech, thus confirming again that many of the early measurements had been wrongly interpreted as being indicative of homogeneous nucleation. Their results also showed that

low lattice misfit is a necessary but not sufficient condition for efficient catalysis, and that chemical effects are often more important. Using the same technique to study the heterogeneous nucleation of copper, Boswell and Chadwick (1980) interpreted their results in terms of eqn. (51.6) with experimental pre-exponential factors of  $10^{31} \text{ m}^{-2} \text{ s}^{-1}$  for the nucleation of copper on a Bi/Sn matrix and  $10^{27} \text{ m}^{-2} \text{ s}^{-1}$  on a Zn/Sn or an Al/Sn matrix, in excellent agreement with a theoretical estimate of  $10^{32} \text{ m}^{-2} \text{ s}^{-1}$ . In view of the later results of Cantor and his collaborators, this agreement may have been fortuitous.

In a series of recent investigations using the embedded droplet method, Cantor and co-workers have shown that experiments of this kind can be improved by using monotectic alloys prepared by rapid solidification. Their studies include the solidification behaviour of lead, cadmium, indium and tin embedded in aluminium-rich matrices (Moore *et al.*, 1990; Zhang and Cantor, 1990abc; Kim and Cantor, 1991), and of lead in copper-rich (Kim and Cantor, 1992) and zinc-rich (Goswami *et al.*, 1992) matrices. Moore *et al.* found that both chill-cast and melt-spun alloys had a bimodal distribution of faceted lead particles with sizes of 1–2 and 5–50  $\mu\text{m}$  in the chill-cast alloys and 5–10 and 50–100 nm in the melt-spun alloys. The bimodal distribution is a result of the formation of the larger particles during cooling through the region where two liquids are immiscible, and the smaller particles during the monotectic solidification of the aluminium matrix. The particles had a cube–cube orientation relationship with the matrix, and a truncated octahedral shape, displaying  $\{111\}$  and  $\{100\}$  facets. The surface energy of a  $\{100\}$  facet was estimated to be about 14% greater than that of a  $\{111\}$  facet, and the solid is thus believed to nucleate on exposed  $\{111\}$  facets of the matrix.

Ternary additions of manganese, copper, zinc and iron did not influence the lead solidification behaviour but silicon was found to be a very efficient catalyst. Two further examples of chemical influence are the effects of magnesium on the nucleation of cadmium on an aluminium matrix and of germanium on the nucleation of lead, also on an aluminium-rich matrix. A comparison of the results from a binary Al–4.5 wt.% Cd alloy with those from a ternary Al–9 wt.% Cd 1.5 wt.% Mg alloy shows that the misfit of the  $\{111\}$  planes is increased by the addition of magnesium but the supercooling also increases, thus demonstrating the importance of chemical effects on the relative catalytic efficiencies of interphase interfaces.

Zhang and Cantor (1992) investigated melt-spun Al–5 wt.% Pb alloys with ternary additions of 0–2 wt.% Ge. The microstructure up to  $\sim 1$  wt.% Ge contained pure lead particles of 5–150 nm diameter. The particles had truncated octahedral shapes and a cube–cube orientation relation with the aluminium matrix which formed on an Al[Ge] solid solution. As the germanium content increases, the misfit decreases but the supercooling to initiate freezing nevertheless increases.

An analysis of these recent experimental results by Kim and Cantor (1994) shows that many of them do not fit eqn. (51.7) at all well. Reasonable values of the pre-exponential factor are obtained only for the nucleation of cadmium and tin on an aluminium matrix; the results for the nucleation of lead on aluminium-, copper- or zinc-rich matrices and those for the nucleation of indium on aluminium are all much lower than the theoretical

value. Kim and Cantor emphasize this by calculating the number of nucleating sites per particle on the assumption that the other components of the pre-exponential term of eqn. (51.7) must have their calculated values. This number ranges from  $10^{-12}$  to  $10^{-7}$  for the four cases above, and this is obviously not physically compatible with the assumptions of the model. The derived number of sites becomes increasingly unrealistic as the deduced contact angle  $\theta$  of eqn. (51.5) decreases from  $27^\circ$  for indium on aluminium to  $4^\circ$  for lead on copper, and it should be noted that the model of Fig. 10.9 is also unrealistic for  $\theta$  values below about  $20^\circ$  when the spherical cap nucleus is only a few atom layers thick. This situation is similar to that encountered in nucleation from the vapour phase and requires a different model of heterogeneous nucleation.

### 65. NUCLEATION OF THE LIQUID-SOLID CHANGE IN ALLOYS

The theory of Chapter 10 did not include transformations which produce a product phase with a changed chemical composition, a complication which is obviously very important in most metallurgical assemblies. The simplest examples of such transformations in fluid phases, e.g. the condensation of a mixed vapour phase or the separation of a single liquid solution into two saturated liquid solutions, have not been much discussed in relation to metallurgical systems and will not be considered here. The nucleation of solid from liquid in a binary alloy, however, is a topic of great importance and is addressed in this section.

Consider the formation of a single solid phase in a binary system in which there is complete mutual solubility in the liquid phase. At a temperature where some alloys are liquid and others solid, the free energy versus composition curves will be of the general form shown in Fig. 14.1. A liquid alloy of composition  $x$  has a free energy per atom represented by the point  $P$ , and if this partially solidifies to a solid solution of composition  $x^s$ , the free energy per atom falls to point  $Q$ . The free energy change is thus  $-(PQ)$  per atom of alloy or  $-(RS)$  per atom in the solid phase as the phases are in the ratio solid:liquid =  $LQ:QS$ . This is the net driving force for the whole reaction at this temperature, but it is important to realize that it is not the driving force for nucleation of the solid. When a very small solid region of composition  $x^s$  has been formed, the liquid will be essentially unchanged in composition. The change in free energy per atom of the nucleus will thus be given by  $-(TS)$ , where  $PT$  is the tangent to the liquid free energy curve at  $P$ . The initial driving force for nucleation of solid of any other composition is given by the same construction; thus  $-(T'S')$  is the change in free energy per atom of an initial solid nucleus of composition  $x'$ .

Suppose the slope of the solid free energy versus composition curve at some composition  $x'$  is algebraically greater than the slope of the tangent  $TP$ . The driving force for nucleation of solid of composition  $x' + dx$  will then be smaller than that for nucleation of solid  $x'$  if  $dx$  is positive, and larger if  $dx$  is negative. If the slope of the free energy curve at  $x'$  is algebraically smaller than that of  $TP$ , the opposite conclusion is

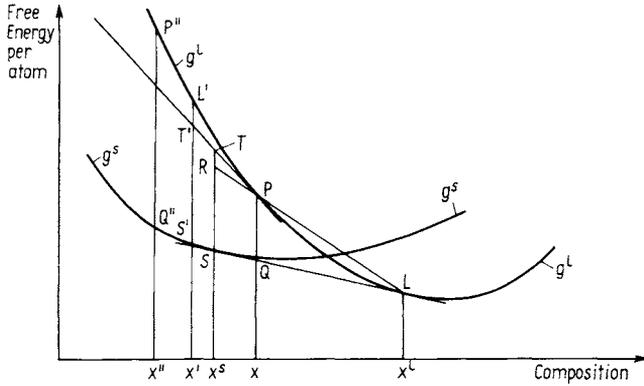


FIG. 14.1. Free energy versus composition curves for liquid and solid phases. The equilibrium free energy change per atom for partial solidification of liquid of composition  $x$  is  $PQ$  and that for composition  $x''$  is  $P''Q''$ . The driving forces for nucleation of solid of compositions  $x^s$  and  $x'$  from liquid of composition  $x$  are  $TS$  and  $T'S'$  respectively.

reached. It follows that the maximum driving force corresponds to the solid composition where the slopes are equal, given by the condition

$$(\partial g^s / \partial x)_{x'} = (\partial g^l / \partial x)_x \quad (65.1)$$

If the composition  $x$  is between  $x^s$  and  $x^l$ , as in Fig. 14.1, the tangent to the liquid free energy curve at  $P$  necessarily has a larger (algebraic) slope than the common tangent to both curves when  $x < x^s$ , and a slope smaller than that of the common tangent when  $x > x^s$ . The composition  $x'$  giving maximum driving force for nucleation is thus greater than  $x^s$  if  $x^s > x^l$ , and is smaller than  $x^s$  when  $x^s < x^l$ . In both cases, the composition of the solid differs from that of the liquid by a greater amount than does the composition of the equilibrium solid. If there is a continuous solid solution with liquidus and solidus temperatures all intermediate between the melting points of pure  $A$  and pure  $B$ , the driving force will always favour initial nucleation of solid richer in the higher melting component. For a continuous solid solution with a (coincident) minimum in the liquidus and the solidus, the driving force will favour the nucleation of solid richer in  $A$  than the equilibrium composition if the overall liquid composition is on the  $A$  side of the minimum, and richer in  $B$  if  $x'$  is on the  $B$  side of the minimum.<sup>†</sup>

The above considerations apply only to nuclei formed at temperatures where the equilibrium state is two-phase. If a liquid alloy of composition  $x''$  (Fig. 10.2) is supercooled below its solidus to a temperature where the equilibrium state is a single homogeneous solid solution of the same composition, the vertical difference between the solid free energy versus composition curve and the tangent at  $x''$  to the liquid free energy versus composition

<sup>†</sup>This conclusion depends on the assumption that both free energy curves have simple U-shapes, so that their slopes increase continuously as  $x$  varies from zero to one.

curve again specifies the variation of the driving force for nucleation with the composition of the solid nucleus. However, because the tangent at  $P''$  may have either a greater or a smaller slope than that at  $Q''$ , it is not possible to generalize on whether this driving force will favour nuclei richer or poorer in A at the equilibrium composition  $x''$ .

An expression is now needed for the rate at which solid nuclei of given composition will form from the liquid melt. Let  $g^l(x', x)$  be the driving force for nucleation of a solid of composition  $x'$  from a liquid of composition  $x$ . The negative of this change  $\Delta g^{ls}(x, x')$  is the free energy change per atom of the nucleus, and is given by

$$\Delta g^{ls}(x, x') = g^s(x') - g^l(x) - (x' - x)(\partial g^l / \partial x)_x \quad (65.2)$$

The formation of a solid embryo of  $n$  atoms having composition  $x'$  will produce a net free energy change

$$\Delta G = n\Delta g^{ls}(x, x') + \eta n^{2/3} \sigma(x, x') \quad (65.3)$$

where the interfacial free energy  $\sigma$  is also a function of the compositions of both liquid and solid. The variation of  $\Delta G$  with  $n$  at fixed  $x'$  has a maximum value which may loosely be called the free energy of activation for the nucleation of solid of composition  $x'$ . To find the composition of the most rapidly forming nuclei, the variation of this critical free energy with  $x'$  must be considered. The two conditions

$$(\partial \Delta G / \partial n)_{x'} = 0 \quad \text{and} \quad (\partial \Delta G / \partial x')_n = 0 \quad (65.4)$$

define the size and composition of the critical nucleus which corresponds to the saddle point in the free energy field, and the critical increase in free energy represented by this saddle point determines the operative nucleation rate in the usual way.

The variation of the surface free energy with composition arises from the so-called chemical part of the interfacial energy, which may be estimated from the nearest neighbour model in terms of the interaction potential  $\Xi$ . For the similar problem of a solid-solid interface, Becker (1937, 1938) assumed a cubic shape for the nucleus and obtained by counting bonds

$$\sigma = \Xi(x' - x)^2 / \rho \quad (65.5)$$

Although the functional dependence on the composition may be much more complex than is indicated by eqn. (65.5), it is nevertheless expected that  $\sigma$  will increase with increasing values of  $|x' - x|$ . Such a variation of interfacial energy with composition tends to reduce  $|x' - x|$  and thus acts in opposition to the driving force inasmuch as nuclei with compositions between the equilibrium and the liquid compositions are preferred. The effects thus partly cancel each other, and the assumption which is often made that the nucleus has the equilibrium composition may be a reasonable working hypothesis. There is then only one variable in eqn. (65.3), and the critical free energy [see eqn. (49.4)] is

$$\Delta G_c = 4\eta^3 \sigma^3 / 27 \{ \Delta g^{sl}(x^s, x) \}^2 \quad (65.6)$$

An analogous development to that given in Chapter 10 now leads to an expression of the usual form for the rate of nucleation in terms of  $\Delta G_c$ .

It is sometimes supposed that fluctuations in composition within the liquid must precede the formation of a solid nucleus. Referring again to Fig. 14.1, it is apparent that with free energy versus composition curves of this form, any fluctuation in which a small group or cluster containing  $n$  atoms temporarily has a mean composition  $x'$  will produce an increase in free energy of magnitude  $L'T'$  per atom in the group; any interfacial energy between the selected group and the remainder of the liquid is ignored. In Borelius' description of solid-state precipitation, the energy  $L'T'$  was considered to be an additional "thermodynamic barrier" to nucleation. However, this reasoning is fallacious; if the nucleation path is liquid ( $x$ )  $\rightarrow$  liquid ( $x'$ )  $\rightarrow$  solid ( $x'$ ), any additional energy required for the first step is regained in the second step.

A detailed comparison of experimental results with nucleation theory for the solidification of a new solid phase from a liquid alloy would require detailed knowledge of the free energy versus composition curves for the liquid and solid phases and of the variation of the interfacial free energy with composition and temperature, in addition to careful measurements of the kinetics of the reaction. As in the case of pure metals, however, some information may be obtained from measurements of a single parameter, the temperature of the onset of rapid nucleation in a large number of individual liquid droplets, or in an aggregate of such droplets.

The simplest type of binary equilibrium diagram is one in which the two components are completely soluble in each other in both liquid and solid states. A typical example is the equilibrium diagram for copper-nickel alloys shown in Fig. 14.2, and an investigation of solidification in these alloys was made by Cech and Turnbull (1951) who used the individual small droplet technique described previously. Their results are shown in

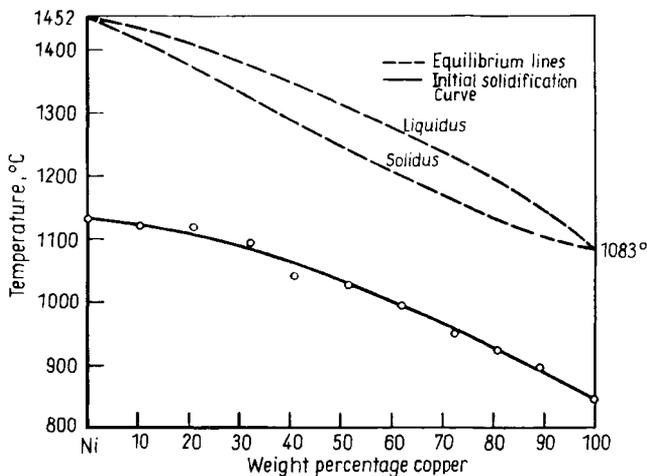


FIG. 14.2. Solidification temperatures of copper-nickel alloy droplets as a function of composition (after Cech and Turnbull, 1951).

Fig. 14.2, the solid line in this figure representing the temperature at which most of the drops solidified. This line is roughly parallel to both liquidus and solidus lines and was originally believed to represent homogeneous nucleation, especially as  $\Delta T^-/T^{\text{sl}}$  had approximately the value of 0.18 found in the early work on pure metals. Similar but slightly smaller supercooling was found by Tarshis *et al.* (1971) in large melts with up to 30% copper. In view of the recent results on pure or nearly pure metals described in Section 64, it now seems very improbable that the measured temperatures of solidification in Fig. 14.2 correspond to homogeneous nucleation.

The contrast between the results of Fig. 14.2 and a usual textbook description of the solidification of a liquid alloy should perhaps be emphasized. It is often stated that freezing begins immediately below the liquidus temperature, and the solid which forms first has the composition given by the solidus at this temperature. As freezing proceeds, the overall compositions of both solid and liquid phases change to the values required by the equilibrium solidus and liquidus respectively, until finally the last drops of liquid solidify and the whole solid has the composition of the original liquid. Clearly this is a description of a process in which equilibrium is maintained at every stage, and it should be approximately valid under conditions of efficient heterogeneous nucleation and rapid solid-state diffusion. Usually, however, the slowness of diffusion prevents equilibrium being maintained in the solid, which then has a cored dendritic structure with spatially varying composition.

This description is clearly not valid if nucleation of the solid (either homogeneously or heterogeneously) requires appreciable undercooling, as in Fig. 14.2. The detailed quantitative information about the free energy versus composition curves required if the initial nucleus composition is to be calculated from eqn. (65.1) is now becoming available from computer simulations, but Fig. 14.2 presents a simpler problem because the nucleation occurs at temperatures well below the solidus, and it is probable that the solid will nucleate with the same composition as the liquid from which it forms. Actually, microscopical examination of Turnbull and Cech's solidified droplets showed that they were not homogeneous but had a cored dendritic structure. This may indicate that the latent heat released by the initial freezing raised the droplet temperature into the two-phase liquid plus solid region before solidification was completed.

In a simple eutectic system, the conditions are more complex. Consider the solidification of an alloy of composition  $x$  in such a system (Fig. 14.3). Crystals of the  $\alpha$  phase may not be nucleated until the liquid has been supercooled appreciably, say to temperature  $T_1$ . These nuclei will grow into larger crystals until some kind of two-phase equilibrium is attained either at temperature  $T_1$  (if the released latent heat can be removed from the assembly) or at some higher temperature determined by the thermal conditions of the experiment. If  $T_1$  is above the eutectic temperature,  $T^{\text{E}}$ ,  $\beta$  crystals cannot form without further cooling, and if the temperature is lowered reasonably slowly, the compositions of the coexisting liquid and solid phases will follow approximately the liquidus and solidus lines and, when the temperature reaches  $T^{\text{E}}$ , the remaining liquid will have approximately the composition of the eutectic mixture,  $x^{\text{E}}$ . The continued growth of the  $\alpha$  phase may result in a dendritic network, subdividing the liquid.

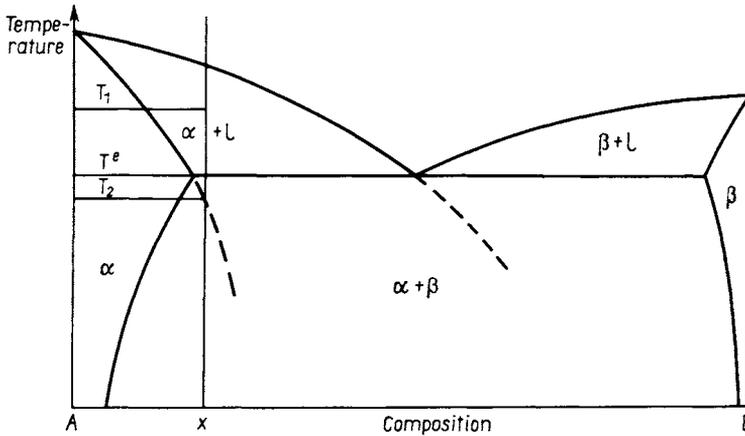


FIG. 14.3. The equilibrium diagram for a simple eutectic assembly.

On reaching the eutectic temperature, further solidification depends on the effectiveness of the existing  $\alpha$  crystals as catalysts for the formation of solid  $\beta$  nuclei. In an extreme case, nucleation of  $\beta$  on  $\alpha$  will require no additional free energy, and the whole of the remaining liquid can then solidify at  $T^E$ . More usually, some supercooling below  $T^E$  will be required to nucleate  $\beta$ . If cooling is slow enough, the compositions of liquid and solid will continue approximately along the metastable extensions to the  $\alpha$  liquidus and  $\alpha$  solidus lines (shown as broken lines in Fig. 14.3). When  $\beta$  crystals eventually nucleate, either heterogeneously or homogeneously, they will do so either with the equilibrium composition or with some other composition richer in B than in liquid, as already discussed. It follows that the liquid close to a newly formed  $\beta$  crystal will have a composition near to that of the metastable  $\beta$  liquidus at that temperature, and this will usually lead to the nucleation of fresh  $\alpha$  crystals. As both  $\alpha$  and  $\beta$  crystals are now able to nucleate, the remainder of the liquid will solidify rapidly as an intimate mixture of small  $\alpha + \beta$  crystals to form a characteristic eutectic structure. The latent heat released may be sufficient to raise the temperature to the eutectic temperature  $T^E$ .

The formation of a eutectic mixture requires the nucleation of both  $\alpha$  and  $\beta$  crystals before solidification is completed. It is possible that an alloy of composition x will solidify completely as metastable  $\alpha$ , even though the equilibrium structure contains  $\beta$  phase. This can happen if the temperature  $T_2$  defined by the metastable  $\alpha$  solidus at composition x (see Fig. 14.3) is above the temperature at which the nucleation rate of  $\beta$  becomes appreciable. The equilibrium  $\alpha + \beta$  structure can then be formed only by solid-state precipitation of the  $\beta$  phase from the supersaturated  $\alpha$  solution.

Similar arguments apply if the temperature  $T_1$  at which the  $\alpha$  crystals first nucleate is below  $T^E$ . If the  $\alpha$  crystals can catalyse the nucleation of  $\beta$  at or near this temperature, or if  $\beta$  crystal can nucleate independently, the whole alloy may solidify with an apparent eutectic microstructure, even though the phases cannot have the equilibrium eutectic compositions. Equivalent results are obtained if the  $\beta$  phase nucleates first and then

catalyses the nucleation of  $\alpha$ . However, if the temperature of first nucleation is below  $T^E$ , and only one phase nucleates, further cooling will lead to the growth of primary crystals of this phase, as described above, followed by a fine, quasi-eutectic mixture.

In the neighbourhood of the eutectic point, either alloy may nucleate first, and the microstructure may be very sensitive to small quantities of impurities which affect the relative nucleation rates. Some important commercial procedures, in particular the "modification" of aluminium-silicon alloys by the addition of very small quantities of sodium to the melt, were formerly attributed to this effect. The modification results in considerable grain refinement, and it was believed to be caused by the suppression of the formation of silicon nuclei, so that nuclei of the aluminium solid solution form first, even in hypereutectic alloys. Later work (Day and Hellawell, 1968; Day, 1970) has provided strong evidence that, in this particular case, the modification is due to a change in the growth characteristics rather than to the nucleation behaviour, but a nucleation effect could, in principle, exist in other systems.

One important conclusion from the above description is that the volume fraction of the two phases in a solidified eutectic mixture will only correspond to that calculated from the equilibrium diagram if one phase acts as a perfect nucleation catalyst for the other.

Hollomon and Turnbull (1951) used the small droplet technique in an investigation of the supercooling of lead-tin alloys. The results are shown in Fig. 14.4, the crosses and squares indicating the temperatures of maximum supercooling of small droplets with various liquid compositions. Alloys with 20–60 at.% Sn also underwent a second marked change in surface structure (observed visually under a microscope) and these temperatures are shown by triangles in Fig. 14.4. In further experiments, it was found that this second critical temperature for the alloys with 20 and 30 at.% Sn was produced by segregation, and was not observed when the solidified alloys were first held at 220 C for some hours.

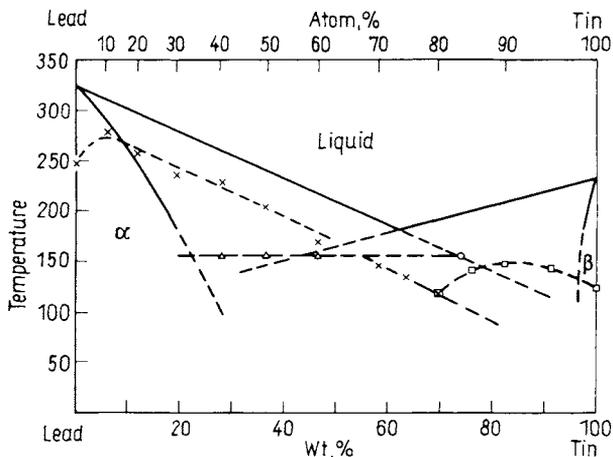


FIG. 14.4. Solidification temperatures of lead-tin alloy droplets as a function of composition (after Hollomon and Turnbull, 1951). The equilibrium phase boundaries are extrapolated below the eutectic temperature.

The results show that the critical temperatures for nucleation are not now parallel to the liquidus and solidus, but increase for small additions of either component to the other. This may reasonably be attributed to an initial decrease in interfacial free energy with composition. Figure 14.4 shows the equilibrium  $\alpha$  and  $\beta$  liquidus and solidus lines, and their metastable extensions below  $T^E$ . The broken line joining the points marked by crosses gives the temperatures of easy nucleation of the lead-rich solid solution. In the 10 at.% Sn alloy, the temperature of such nucleation is just below the solidus, and the whole alloy solidifies when nucleation begins. In the 20 and 30 at.% alloys, nucleation leads to the formation of a lead-rich  $\alpha$  phase and liquid. If the cooling is slow, these alloys also solidify entirely as  $\alpha$  phase when the solidus temperature is reached. With more rapid cooling, the rate of diffusion is inadequate to maintain the equilibrium  $\alpha$  and liquid compositions and metastable  $\beta$  is formed, as in the 40 and 50 at.% alloys. These alloys form  $\alpha$  phase and liquid below the temperature of  $\alpha$  nucleation, and the compositions change approximately along the metastable solidus and liquidus lines until the temperature indicated by triangles is reached. At this temperature, liquid of the composition marked by a circle on the metastable  $\alpha$  liquidus is sufficiently supercooled to produce  $\beta$  nuclei, and both  $\alpha$  and  $\beta$  phases then grow. The triangles thus represent the temperature of  $\beta$  nucleation after  $\alpha$  crystals have formed, and as  $\beta$  then always nucleates from liquid of the same composition, the triangles are all at the same temperature.

The alloy with 90 at.% Sn solidifies at a single temperature lying above the  $\alpha$  liquidus line. This must represent the temperature at which primary  $\beta$  phase nucleates and, after growth of these nuclei, the remaining liquid will have a composition on the extended  $\beta$  liquidus. For the liquid of this composition, the temperature of primary  $\beta$  nucleation is well below that of primary  $\alpha$  nucleation, so that growth of the primary  $\beta$  leads immediately to nucleation of  $\alpha$ , and the whole alloy solidifies. The line of squares on the right of Fig. 14.4 represents the temperatures at which primary  $\beta$  forms. For the composition of the circle point, the nucleating temperature of primary  $\beta$  (shown by the square) is about 10°C below the temperature at which  $\beta$  nucleates in the presence of  $\alpha$  (shown by the triangles). Although this difference may represent a mild catalytic effect of the  $\alpha$  phase in promoting  $\beta$  nucleation, it is almost within the experimental error.

Another small droplet technique used by Southin and Chadwick (1978) is to anneal the specimen in the partly liquid state above the three-phase equilibrium temperature until repeated slow cooling curves give reproducible results. In most cases, at least two thermal arrests were observed on slow cooling, one close to the eutectic temperature and one (or more than one) at a considerably lower temperature. Nucleation close to the eutectic temperature was ignored as being possibly due to contamination at grain boundaries, so that only the lower arrests representing nucleation within liquid droplets enclosed by solid grains ("internal" droplets) were considered. The results showed undercoolings below the eutectic temperature varying from 2 K to 115 K, many of which were much larger than previously reported values.

Another important solidification sequence takes place when there is a peritectic reaction. In Fig. 14.5, an alloy of composition  $x$  may be supposed to nucleate  $\alpha$  crystals at

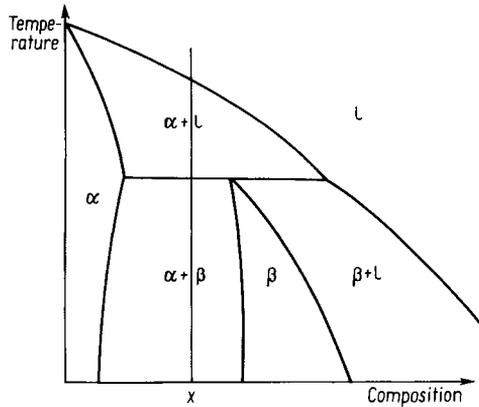


FIG. 14.5. Diagram illustrating a peritectic reaction.

a temperature just below the true  $\alpha$  liquidus and, with slow cooling, the compositions of the  $\alpha$  phase and the liquid move along the solidus and liquidus until the peritectic temperature is reached. At this temperature there is, in principle at least, a reaction between solid  $\alpha$  and the liquid to yield a product of solid  $\beta$ . The reaction will actually begin at the peritectic temperature if the  $\alpha$  phase catalyses the nucleation of the  $\beta$  phase, or if  $\beta$  nucleation is otherwise possible, but it is important to realize that the reaction cannot begin until  $\beta$  is nucleated. If there is no nucleation of  $\beta$  at or near the peritectic temperature, the solidification will continue to follow the metastable liquidus and solidus temperatures until either all of the alloy forms metastable  $\alpha$  or nucleation of  $\beta$  begins. In a peritectic reaction, however, nucleation of both phases does not necessarily result in the solidification of the whole alloy without further lowering of the temperature. The primary  $\alpha$  crystals eventually become coated with a layer of  $\beta$ , and the later stages of the reaction require diffusion through these  $\beta$  layers. With a sufficiently thick coating, reaction may become very slow, and further cooling may be required to complete the solidification. The resultant non-equilibrium microstructures show peritectic "sheaths" of  $\beta$  around  $\alpha$  particles.

The course of the peritectic reaction will clearly differ from the above description if the  $\alpha$  phase fails to nucleate above the peritectic temperature or when alloys with compositions to the right of the peritectic point in Fig. 14.5 are considered. Experimental investigations of the microstructures produced by peritectic reactions (Uhlmann and Chadwick, 1961) are in general agreement with the above description.

Monotectic reactions have been mentioned above as providing ideal conditions for the embedded drop technique, but no systematic study of the nucleation behaviour appears to have been reported. In a monotectic reaction, a single liquid phase decomposes on cooling and forms a solid phase and another liquid phase. Livingston and Cline (1969) investigated a monotectic reaction in copper-lead alloys, and showed that the microstructure may be very sensitive to the growth conditions.

It remains to discuss solidification when the solid phase is neither a pure metal nor a primary solid solution but is an intermetallic compound or an intermediate phase of varying composition. There is no sharp distinction between the two categories, but intermetallic compounds may usefully be distinguished by the extremely restricted range of composition over which a single-phase field is stable, and by the fact that they usually are stable in the solid form up to quite high temperatures, so that the equilibrium liquidus and solidus lines have a coincident maximum, often at some stoichiometric composition  $A_xB_y$ , where  $x$  and  $y$  are integers. In extreme cases, intermetallic compounds occur only at some fixed composition, and are then represented as vertical lines on an equilibrium phase diagram.

Other intermediate phases are quite different, often appearing as solid solutions with quite wide variations in composition, and often forming from the liquid by means of a peritectic reaction, rather than melting and freezing congruently as do many intermetallic compounds. For such an intermediate solid solution, the conditions defining the kinetics of solidification are exactly equivalent to those described above for primary solid solutions, and need no further discussion. In principle, the same remark applies to intermediate phases, but the very strong variation of free energy with composition makes the composition of maximum driving force constant at the composition of the stoichiometric alloy.

## 66. THE GROWTH OF CRYSTALS FROM THE MELT

In this section, the way in which an existing solid nucleus grows into the surrounding liquid is considered. When a crystal is produced from the vapour phase or from a dilute liquid solution, the rate of growth is usually limited either by the rate at which individual atoms or molecules can reach the vicinity of the crystal surface, or by the rate at which they can attach themselves to the surface. Similar limiting processes may also govern the rate of growth of a solid phase from a liquid alloy but there are no difficulties of supply when a pure metal solidifies and the rate of growth may then be determined mainly either by the net rate at which atoms stick to the solid surface, or by the rate at which the released latent heat of solidification can be removed from the region of the solid-liquid interface. As already emphasized, the rate of heat removal is often the limiting factor when pure metals are solidified and, once nucleated, the solid grows very rapidly if the melt is appreciably supercooled before freezing begins. In a bulk liquid, this rapid solidification usually gives a solid composed of dendritic (tree-like) structures, and the remaining interdendritic liquid solidifies more slowly, as its temperature has been raised by the released latent heat. This dendritic growth is generally equi-axed, and the released heat is conducted away through the cooler liquid ahead of the interfaces. For a positive growth direction, the temperature gradient in the liquid near to the interface is thus negative, whilst that in the solid is nearly zero (see Fig. 14.6).

Many experiments on solidification processes do not correspond to the free growth conditions just described, but involve the external imposition of selected rate of growth.

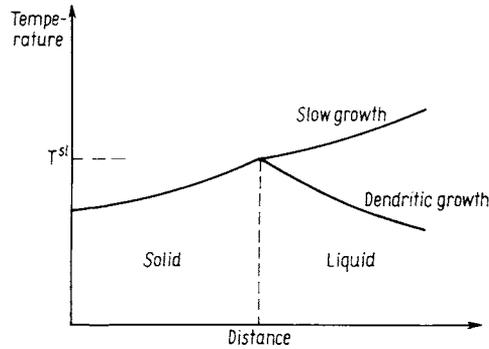


FIG. 14.6. Temperature distribution in liquid and solid for the slow growth of a pure metal from its melt.

These experiments utilize directional solidification in which the melt is cooled in an externally imposed temperature gradient. The temperature gradient in the liquid ahead of the interface is now positive and the heat has to be removed through the growing solid (see Fig. 14.6). Directional freezing experiments of this type were pioneered by Chalmers and his collaborators, and they have provided much of the available experimental information on freezing processes. In such experiments, the growth rate,  $\dot{\gamma}$ , and the temperature gradient at the interface,  $G$ , are externally controlled process variables, whereas in growth in an undercooled melt only the initial supercooling can be regarded as externally imposed.

The presence of additional components in solution in the liquid melt, even in quite small concentrations, may very significantly change the way in which a single-phase solid grows. The most obvious effect of a second component is to produce an unevenness of composition in the solid solution which forms from the liquid. This normally arises because at temperatures between the liquidus and the solidus, the compositions of the solid and liquid phases in equilibrium with each other are not identical. There is thus partitioning of solute during growth and a diffusion field develops ahead of the interface, so that the rate of solute diffusion, rather than the heat flow, may be the most important factor determining the growth rate. This section is primarily concerned with the solidification of pure metals, and the solidification of alloys is examined in Section 67. However, before discussing more detailed growth theories, it is instructive to consider some relationships between characteristic linear features of the solidified microstructures of both pure metals and alloys and the physical phenomena which produce these features. The resultant scaling relations are due mainly to Trivedi and Kurz (1994ab).

A single-phase solid growing into a liquid may have various different macroscopic morphologies which depend on externally imposed conditions. Thus the interface between a pure metal and its melt may be either planar (or slightly curved) or it may be dendritic. Dendrites frequently grow in arrays, especially in unidirectional solidification, but in alloys there is another type of interface which is described as “cellular”. A cellular

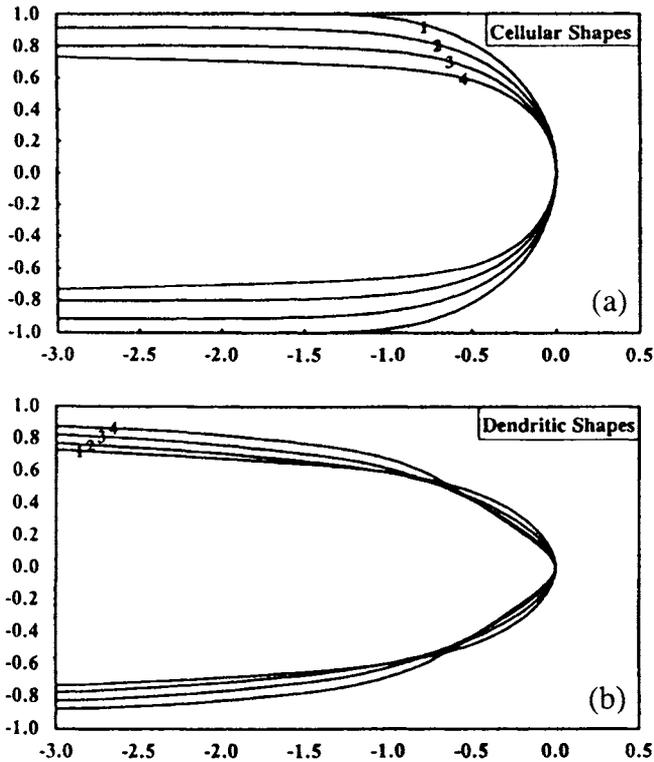


FIG. 14.7. Schematic shapes of (a) cellular structure; (b) tips of dendrites (after Lu and Hunt 1992).

structure is a periodic arrangement of protruding regions (pseudodendrites), separated by deep liquid valleys; it is essentially similar to a dendritic structure without side-arms, but the leading parts of the interface protrusions are flatter than the near paraboloidal shapes of primary dendritic tips (see Fig. 14.7). Both cellular and dendritic structures in alloys are stabilized by differences in composition between the tip regions and the interdendritic or cell boundary regions. The cell or dendritic structure thus controls the distance over which appreciable variations in solute content are found in the inhomogeneous ("cored") solid solution.

As the imposed growth rate in the unidirectional solidification of an alloy is increased, the inherently unstable planar interface breaks down to form a cellular structure and, at still higher growth rates, the interface becomes dendritic. This was once considered to be the whole story, but recent work has shown that, if the velocity continues to increase, the reverse transitions, dendritic  $\rightarrow$  cellular  $\rightarrow$  planar or, in some conditions, dendritic  $\rightarrow$  planar, will take place, so that in general a planar interface may be observed at either very slow or very fast growth rates.

The structure of a dendrite in a cubic crystal is shown schematically in Fig. 14.8. It grows initially as a primary stalk or trunk from which secondary, tertiary and higher-order

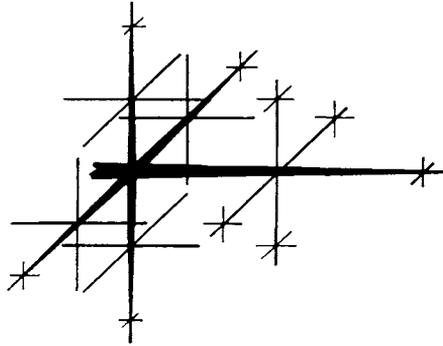


FIG. 14.8. The structure of a dendrite (schematic) (after Weinberg and Chalmers, 1951).

branches develop at a later stage. In cubic crystals, the dendritic arms are all mutually perpendicular and correspond to  $(100)$  directions of the lattice. A complete theory of this structure will have to deal with the dependence on the growth conditions of various characteristic microstructural lengths, among which are the radius of the tip of the primary dendrites, their spacing normal to the growth direction (for unidirectionally frozen materials), the spacing and tip radii of secondary, tertiary, etc. arms and the distance from each tip to the first subsidiary arm to form. This is a task of impossible complexity, especially if the same theory has also to account for the interface transitions and to predict the characteristic lengths (cell spacing, depth and width of intercell "grooves", etc.) which define the cellular microstructure. Nevertheless, Trivedi and Kurz (1994ab) have pointed out that much useful information about the way in which such microstructural lengths depend on process variables may be obtained by representing the main physical phenomena in terms of characteristic lengths which are then related to features of the microstructure.

The important physical processes governing growth are solute diffusion (in alloys only), thermal diffusion, capillarity (i.e. the Gibbs–Thomson effect of interface curvature on phase equilibrium), and possibly convection, gravitational forces and interface attachment kinetics. Convection is often of great importance in industrial practice, but is usually considered to be an unnecessary complication in laboratory experiments where its effects may be minimized by upward unidirectional solidification in a vertical furnace. Equally, gravitational segregation can be minimized by a suitable choice of alloy system, or by experiments in space, and interface kinetics require normally a very small driving force and hence produce negligible additional supercooling at the interface. Their effects are thus frequently neglected.

The remaining physical phenomena may be associated with three length scales with very different typical values. Solute diffusion (if required) and thermal diffusion may both be described in terms of effective "diffusion lengths"  $l_D = D^l/\gamma$  and  $l_T = D^l/\gamma$  for growth into an undercooled pure metal or alloy melt, where  $D^l$  and  $D^l_\gamma$  are solute and

thermal diffusion coefficients respectively and  $D_T^l$  is given in terms of the thermal conductivity  $K_T^l$  and the atomic specific heat of the liquid  $C_p^l$  by

$$D_T^l = K_T^l v^l / C_p^l \quad (66.1)$$

For the unidirectional solidification of an alloy,  $l_T = \Delta T_0 / G$ , where  $\Delta T_0$  is the difference in the equilibrium liquidus and solidus temperatures. Curvature (capillarity or surface free energy) effects have a characteristic length given by  $l_r = \Gamma' / Q'$ , where  $Q'$  is the interfacial free energy per unit area divided by the entropy of transformation per unit volume, and for a pure metal is the latent heat divided by the specific heat. The "capillarity length"  $l_r$  for a pure metal is one-half the radius of a critical nucleus, as given by eqn. (46.12). Typical values for these scaling lengths have been listed by Trivedi and Kurz; for growth rates of  $10^{-6}$  and  $10^{-4} \text{ m s}^{-1}$ , and associated gradients of  $10^2$  and  $10^4 \text{ K m}^{-1}$  and  $\Delta T_0 = 10 \text{ K}$ ,  $l_D$  varies from 5 mm to 50  $\mu\text{m}$ ,  $l_T$  in unidirectional solidification from 100 mm to 1 mm and in a supercooled bath from 50 m to 0.5 m, whilst  $l_r$  is typically 10 nm.

Trivedi and Kurz have shown that, with the above reservations, any microscopic length  $L^*$  (even including eutectic spacing) which results from solidification can be expressed in terms of the three characteristic physical lengths by the general equation

$$L^* = A [l_D]^a [l_T]^b [l_r]^c \quad (66.2)$$

where  $a + b + c = 1$ .

These equations, albeit without many details of the real physics, enable the form of the functional dependence of the microstructural variables to be established either theoretically or experimentally, and also give criteria for the transitions in microstructure which are observed in directional freezing. These are listed as:

- (i) limit of planar interface in slow growth (constitutional supercooling condition; see below):

$$l_D = l_T \quad (66.3a)$$

- (ii) limit of planar growth at high velocities:

$$l_D = i l_r \quad (66.3b)$$

where  $i$  is the (assumed constant) distribution coefficient, i.e. the ratio of the solid composition to the liquid composition with which it is in equilibrium.

- (iii) cell-to-dendrite transition at low velocity:

$$l_D = i l_T \quad (66.3c)$$

and

- (iv) dendrite-to-cell transition at high velocities

$$l_D = \alpha l_r \quad (66.3d)$$

where  $\alpha$  is an unknown constant.

The analysis shows that  $a=b=c=1/3$ , for example, describes the wavelength at which, at a critical velocity, a planar interface becomes unstable; as noted above this also requires  $l_D = l_T$ , so the relation can be expressed by  $a=1/2$ ,  $b=1/2$ ,  $c=0$ . For some conditions, the thermal diffusivity does not influence the microstructure and  $a=c=1/2$ ; this applies when  $L^*$  represents the tip radius of a primary alloy dendrite, or the initial spacing of secondary dendrites, and in general to structures formed close to an isothermal interface. In contrast, there is obviously no solute diffusion in a pure melt which is undercooled, and in this case the microstructural parameters will have dependencies of  $a=0$ ;  $b=c=1/2$ . These scaling laws show the way in which the microstructure depends on processing parameters such as the solute content of the alloy and the interface velocity in directional solidification.

Most detailed theories of dendritic growth are based on the assumption that a steady state has been attained and that there is near equilibrium between the two phases at the growing interface. In the solidification of a pure metal, for example, the two phases are in equilibrium in the immediate vicinity of a planar interface held at the pressure and temperature ( $T_m$ ) of the melting point. A curved interface, convex towards the liquid at the same external pressure, will be in equilibrium with the liquid at a different temperature because of the Gibbs–Thomson effect which will displace the equilibrium temperature by an amount  $\Delta T_r$ . From eqn. (49.13) and the equation for spherical solid particles which corresponds to eqn. (46.12), it follows that the Gibbs–Thomson supercooling may be written:

$$\Delta T_r = 2\sigma^{sl}v^1/r\Delta s^{sl} \quad (66.4)$$

for the symmetrical case when the two principal radii of curvature,  $r_1$  and  $r_2$ , both equal  $r$ , and the same equation applies to other conditions if  $2/r$  is replaced by  $r_1^{-1} + r_2^{-1}$ .

A moving interface must be at a slightly different temperature from this equilibrium temperature of  $T^{sl} - \Delta T_r$  because a finite driving force is needed to displace the interface at a given velocity. This term represents kinetic processes at the interface, and may be written as a supercooling of the interface by an amount  $\Delta T^B$ , so that the curved interface is supercooled by an amount  $\Delta T^-$  from the melting point where

$$\Delta T^- = T^{sl} - T_i = \Delta T_r + \Delta T^B \quad (66.5)$$

and this equation may be compared with eqn. (54.62) which gives the corresponding result for an alloy in terms of the change in composition at the interface required by material diffusion. The kinetic undercooling is usually very small (of the order of  $10^{-1}$  K for a growth rate of  $10^{-1}$  m s $^{-1}$ ) so that it is frequently neglected in theories of growth.

In the remainder of this section, the growth of pure metals will be considered in more detail. Non-dendritic growth of a pure material may be observed under conditions of slow freezing at a temperature just below its melting point; such conditions may be obtained, for example, by directional solidification in which case the interface is nearly at the melting temperature, and the liquid is at a higher temperature. If any portion of the interface grows temporarily a little faster than the remainder, so as to form a small projection, it encounters liquid at a higher temperature and its growth ceases. Planar or slightly curved

interfaces are thus stable against fluctuations in shape and the resulting slow growth can be controlled externally by the rate of heat withdrawal, as in many methods (see, e.g., Bridgeman, 1924; Czochralski, 1913) of preparing single crystals. In free growth conditions, however, the liquid in the vicinity of the interface is supercooled so that a small projection encounters colder liquid and so grows still more rapidly. A planar interface is thus inherently unstable and dendrites will grow out from it. This type of growth cannot be controlled easily.

The temperature distribution in liquid and solid during free growth and controlled growth in a temperature gradient is shown schematically in Fig. 14.6; the latter condition is sometimes described as a temperature inversion. Slow growth in three dimensions should lead to polyhedral crystals, but it is almost impossible to arrange because of the difficulty of removing the released latent heat. Non-dendritic growth in pure metals is thus studied mainly by controlled unidirectional freezing. When intermetallic compounds are grown from a liquid alloy of non-stoichiometric composition, the crystals are frequently found to have well-developed external faces. This is possible because diffusion in the liquid becomes an important rate-controlling process, and planar interfaces are then stable even if the adjacent liquid is slightly supercooled.

Now consider to what extent do the conclusions of Chapter 13 apply to a crystal face in contact with its own liquid; i.e. does the slow growth in the absence of defects require two-dimensional nucleation of each successive layer? The energy of a surface defect will be lowered by contact with the liquid, and hence the temperature at which surface disordering becomes extensive (often called the surface melting temperature) will also be lowered. This problem was mentioned briefly on p. 167; the main result of the simple two-level Bragg-Williams treatment (Jackson, 1958ab) is that the surface is "disordered" (i.e. more than about one-half of the sites in the outermost layer are unoccupied), if

$$a^{\dagger} = \Delta h \xi / kT \leq 2 \quad (66.6)$$

where  $\Delta h$  is the latent heat for the change from the crystalline phase to the phase with which it is in contact,  $T$  is the surface temperature and  $\xi$  is the fraction of nearest-neighbour bonds which lie in the surface plane. The parameter  $a^{\dagger}$  represents the ratio of the binding energy at the interface to the thermal energy and is thus a natural measure of the amount of disorder at the interface. For solids in contact with their melts,  $a^{\dagger}$  depends mainly on the dimensionless entropy of melting  $\Delta s^{\text{sl}}/k = \Delta h^{\text{sl}}/kT^{\text{sl}}$  and this is close to unity for most metals. The above condition for a disordered ("rough" or "melted") solid-liquid interface is thus satisfied at temperatures approaching  $T^{\text{sl}}$  as, even for close-packed planes,  $\xi < 1$ . A few organic and inorganic compounds have relatively small entropies of melting and hence satisfy the condition (66.6) but, for most molecular crystals,  $\Delta s^{\text{sl}}/k$  exceeds five, and minimum free energy then corresponds to only a small amount of disorder in close-packed surfaces. The liquid-solid surfaces of such crystals are expected to develop facets parallel to close-packed planes. Metalloids and semiconductors such as bismuth, antimony, silicon and germanium have values of  $\Delta s^{\text{sl}}/k$  of between two and three, and the interface structure cannot readily be predicted from this simple theory.

More complex calculations, such as those of Burton and Cabrera described in Section 19, give essentially similar results. A multilayer model which predicts a very disordered and diffuse interface for  $a^l = 0.5$  with part occupancy of atomic sites extending over about 16 layers was later developed for  $\{100\}$  interfaces. As  $a^l$  increases, the interface becomes sharper, and it is essentially confined to two layers for  $a^l = 3$  or 5. Other workers have obtained generally similar results for the dependence of surface roughness on  $a^l$  for a multilevel pair model and from computer simulations using a Monte Carlo (random exchange) procedure.

Frank (1952) pointed out that, in addition to the equilibrium structure of a face, it is necessary to consider the rate at which this equilibrium configuration can be established. As atoms are deposited on the solid, they tend to produce atomically flat surface regions, so that continuous growth at random points is possible only if the rate at which the surface attains its equilibrium disorder is greater than the rate at which this is removed by growth. In the analogous problem of growth from the vapour phase, the low activation energy needed to form a kink in a step implies that there is little difficulty in assuming that a moving step maintains its equilibrium kink concentration. However, it may not be justifiable to assume that disordering of a solid-liquid interface takes place with comparable ease because the process is now cooperational (see p. 150). If the rate of disordering is slower than that of deposition, continual *free* growth must be controlled either by the disordering rate ("interface control") or else must require two-dimensional nucleation or emergent dislocations to provide interface steps.

In practice, it is difficult to determine experimentally whether or not dislocations are required for slow growth, as the experimental conditions which give slow growth do not allow free growth. Chalmers and Martius (1952) pointed out that if the interface in a unidirectional freezing experiment contains re-entrant corners or edges, the constrained heat flow will ensure that these corners or edges cannot be removed as growth proceeds. Thus, unless a close-packed plane is exactly normal to the growth direction, the experimental conditions eliminate the need for either repeated two-dimensional nucleation or growth by the Frank dislocation mechanism.

Many attempts have been made to obtain experimental information on the growth process by examining the surface of a growing crystal which has been suddenly separated from the liquid into which it was growing. Such experiments (see, e.g., Elbaum and Chalmers, 1955) usually reveal a surface consisting of flat regions separated by steps of height 0.1–1  $\mu\text{m}$ , and this has often been called a platelet structure. However, it is not now believed that this structure is characteristic of the liquid-solid interface. Chadwick (1962) and others have concluded that for energetic and hydrodynamic reasons it is almost impossible to prevent a thin film of liquid from adhering to the solid surface, however great the acceleration produced by the splitting device. The platelet structure is thus a result of the solidification of this thin film in contact with the metal vapour or with air, and this interpretation is supported by the observation that similar platelet structures may often be observed on the free surfaces of solidified metal ingots. The anisotropy of the surface free energy of the solid may provide the driving force for the formation of steps of minimum height which then cluster because of kinetic effects such as those discussed in Section 60.

If crystal faces of any orientation are composed of close-packed planes and steps, this has other implications. An example is a suggestion (Chalmers, 1949, 1954) that different crystal faces may have different equilibrium melting temperatures, which is unlikely to be true if the atomic structures of all interfaces are essentially the same. The free energy per atom is then the same for all surfaces; they are all in equilibrium with the same external vapour pressure, and hence with liquid at the same temperature.

Almost all crystals grown from the melt contain imperfections and/or some kind of substructure. The degree of imperfection may be examined by X-ray or  $\gamma$ -ray topography, by optical or electron microscopy and by various diffraction techniques. The existence of different effects has led to some confusion in the nomenclature. The difficulties arise mainly with the term "mosaic structure", introduced by Ewald, which is commonly applied to all crystals giving diffracted beams of high intensity (proportional to the square of the structure factor) and high angular breadth (some minutes of arc). As Darwin (1922) emphasized, ordinary X-ray methods show only that the atomic arrangement in such crystals is not exactly regular over distances greater than a few thousand interatomic spacings. They thus do not distinguish between situations in which the dislocations are mainly concentrated into boundary regions between slightly misoriented crystal blocks and more random distributions in which the dislocations may form a three-dimensional Frank network. The term mosaic is not strictly applicable to structures of the latter type and it is unfortunate that its use is traditionally associated with a relatively uninformative type of X-ray image. Both types of structure are known to exist in practice and, as already mentioned, a few examples of near-perfect crystals (such as type I diamonds) have also been found.

Artificially produced near-perfect crystals of macroscopic size (as distinct from whiskers) were first obtained with very pure semiconducting materials such as silicon and germanium which were grown from the melt using the Czochralski crystal pulling technique in which a solid crystal is slowly withdrawn on a rotating holder from a melt containing liquid at a controlled temperature. Early work using etch pits suggested that regions of crystal as large as  $10 \text{ mm}^2$  may be free of dislocations, and this was supported by the very narrow range of the X-ray reflections. The Czochralski technique was later shown to produce crystals of copper more perfect than those obtained by the Bridgman method, although crystals of very low dislocation density have also been grown by the latter technique. Subsequent development of the Czochralski technique enabled crystals of copper and silver to be grown with virtually zero dislocation densities (Sworn and Brown, 1972; Tanner, 1972, 1973).

X-ray topography in conditions of anomalous transmission is ideally suited to the examination of crystals of copper and silver about 1 mm thick, and shows that large volumes of the crystals are dislocation-free. However, the presence of small dislocation loops formed, for example, by condensation of excess vacancies, would escape detection by this method. Such loops are believed to be responsible for "black-spot defects" seen in electron metallography, and Tanner has noted that they were never found in his highly perfect copper and silver crystals. In order to obtain crystals of this perfection, all the possible ways of forming dislocations from inclusions, occasional impurity atoms, local

plastic deformation caused by thermal stresses, etc. have to be minimized by special techniques such as growing through a fine neck and using very slow pulling speeds. Tanner found that dislocation-free silver could only be grown at speeds much slower than those used for copper, but for both metals there appears to be an inverse relation between the speed and the maximum diameter of dislocation-free growth.

Unless special precautions are taken, most crystals grown from the melt have substructure of the "macromosaic" or "lineage" types. The lineage structure was first discussed by Buerger (1934) who suggested that branches of a crystal growing from a common nucleus are distorted in varying amounts by external influences and internal stress fields. When two distorted branches meet again, they will no longer fit together exactly, and macroscopic surfaces of orientation discontinuity will be formed. Misorientations of this kind will obviously be expected when growth is dendritic as impurity particles, thermal gradients, etc. will cause plastic bending in the individual dendrites, and the misfit must be accommodated by dislocation arrays when dendritic arms join together. However, many crystals grown from the melt by slow solidification also have a substructure of the macromosaic type, and this may be sufficiently coarse to be readily recognized by optical techniques or by the splitting of spots in Laue photographs taken with beams of conventional width.

In the case of a slowly grown single crystal, the macromosaic is a fibre structure in which the sub-boundaries (or "striations") are parallel to the general growth direction and are misoriented relative to each other by small amounts (typically  $1^\circ$ ) about an axis which is parallel to the growth direction for a large number of metals (Teghtsoonian and Chalmers, 1951). This means that the boundaries between the subgrains are mainly small angle tilt boundaries, composed of edge dislocations parallel to the growth axis (see Chapter 8). The cross-sectional geometries of the subgrains differ in different metals. Teghtsoonian and Chalmers (1952) found that in tin the subgrains are approximately square section rods, but the section is wafer-shaped in zinc and is irregular in aluminium.

In a detailed investigation of the substructures in tin, Teghtsoonian and Chalmers found that the average widths of the sub-grains decreased from 1.23 mm to 0.25 mm as the growth rate increased from  $0.02 \text{ mm s}^{-1}$  to  $0.2 \text{ mm s}^{-1}$ , and the average misorientation between neighbouring rods decreased from  $2.8^\circ$  to  $1.2^\circ$ . Successive misorientations tended to alternate in sign so that the maximum deviation from the mean was less than  $4^\circ$ . Crystals were grown with a  $\{001\}$  plane as the top surface, but with various growth directions. The boundaries were always parallel to the growth direction at slow rates of growth but deviated increasingly towards the  $\langle 110 \rangle$  direction for faster rates. Similar experiments with lead crystals with a vertical  $\langle 001 \rangle$  axis showed that the macromosaic boundaries are rotated towards a  $\langle 100 \rangle$  direction. In experiments in which the width of the specimen increased at some stage during growth, new sub-boundaries did not appear immediately but developed within a few centimetres. The misorientation across these boundaries was initially very small, but increased to the same limiting value as that of the original sub-boundaries.

The sub-boundaries or striations are believed to form from aggregates of dislocations, and theories thus depend on assumptions about the formation of these dislocations. The

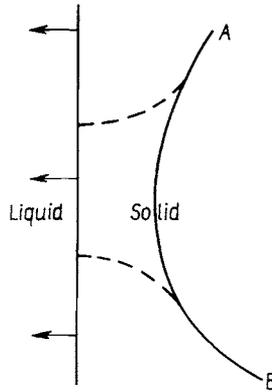


FIG. 14.9. A dislocation line extending into the liquid–solid interface.

increasing misorientation implies an increase in dislocation density along the length of the crystal and a simple possible mechanism is shown in Fig. 14.9, which illustrates an existing dislocation gliding or climbing to the liquid–solid interface, after which it grows with the crystal as two dislocation lines.

Electron microscopic observations on quenched metals have shown that excess vacancies condense to form dislocation loops under appropriate conditions, and Teghtsoonian and Chalmers suggested that their results could be explained qualitatively by the condensation of vacancies in sheets which then condense to form edge dislocation loops. The high concentration of vacancies at the melting point will lead to supersaturation in the cooler regions behind the interface, and the resultant edge dislocations should be sufficiently mobile to assemble into surfaces. At higher growth rates, the smaller time available for vacancy diffusion will lead to smaller subgrain sizes. The observed dependence of boundary direction on growth rate was assumed by Teghtsoonian and Chalmers to be due to a preferred plane for vacancy condensation.

These ideas were developed in more detail by Frank (1956; 1956–7) and a description of his theory was included in the first edition of this book. The results showed that the vacancy supersaturation close to the interface is too low for the direct formation of vacancy discs which intersect the interface, and in order to explain the experimental results, Frank had to assume that loops nucleated behind the interface grow forward sufficiently rapidly by climb to intersect the interface, towards which they are attracted. Schoeck and Tiller (1960), however, calculated that the maximum rate of climb is smaller than typical interface velocities, and they also made an experimental test of the vacancy theory by growing crystals of tin and lead at ambient temperatures only 20 K below their melting points. This should result in a striation-free substructure if the vacancy theory is correct as the supersaturation of vacancies is then quite small. In fact, the crystals showed fully developed substructures of the usual kind.

Other experimental evidence against the vacancy theory has gradually accumulated, and the present view is that the substructures have their origin in accidental inclusions and

impurities and in thermal stresses. As will be described in Section 67, crystals containing appreciable amounts of solute develop a second kind of substructure, the boundaries of which consist of solute-rich or solute-poor regions, and the macromosaic boundaries are then found to be associated with these "cellular" boundaries. In purer materials, cell structure is not encountered, but an investigation by Atwater and Chalmers (1957) showed that one decisive factor in the formation of a macromosaic may be the presence of very small amounts of impurities. They found that the boundaries of the substructure were not straight and were discontinuous when tin of very high purity was used to grow the crystals. The presence of impurities in either tin or lead changed the extent to which the boundaries deviated from the growth direction towards a preferred crystallographic direction. Similar observations were made by Jaffrey and Chadwick (1968).

Despite the frequent existence of a striation substructure, most of the dislocations in a melt-grown crystal are present as a random network, rather than being concentrated into the sub-boundaries. In early work on lithium fluoride crystals, Washburn and Nadeau (1958) showed that the dislocation density ( $10^7$ – $10^9 \text{ m}^{-2}$ ) is significantly increased by mechanical or thermal stresses, but is not appreciably affected by large variations of growth rate or thermal gradient. The authors concluded from these observations that vacancy condensation is unimportant in generating dislocations. The dislocation density in copper single crystals is apparently unaffected by the growth rate, but decreases with increasing purity. The experience which has accumulated in the practical aspects of the growth of highly perfect, single crystals shows that success depends upon attention to purity and thermal stresses, and especially on the use of very good quality seed crystals, any remaining dislocations in which are prevented from growing into the main crystal by devices such as a narrow neck.

One possible model for the production of new dislocations at the liquid–solid interface, as distinct from the multiplying of existing dislocations, involves the entrapment of small impurity particles initially present in the melt. Particles are either pushed ahead of the interface or incorporated into the solid, depending on their grain size, the interface velocity and the relative surface free energies. A particle which is incorporated within the solid may generate dislocations because of the lattice misfit. Jaffrey and Chadwick (1968) considered that this theory provides the most plausible explanation of the striation substructure, and they suggested that the sub-boundaries arise because of preferential sites for incorporation of impurity particles, rather than by agglomeration of random dislocations. The model is consistent with several experimental observations that the inter-striation distance and the incubation distance for the formation of the substructure are both inversely proportional to the velocity of the interface, but it does not adequately explain the increase in tilt angle with distance.

If the vacancy supersaturation does not lead to high dislocation densities, it must be relieved in some other way. When dislocations are present, some of the vacancies may condense on them, causing the dislocation to climb out of the crystal. Use is made of this process when eliminating unwanted dislocations from a seed crystal (Dash, 1958). However, in regions which are locally free of dislocations, the excess vacancies which cannot escape to a free surface must condense either as small planar discs, or as

three-dimensional small tetrahedral or spherical pores. Any resultant dislocation loops which are formed must be very small as they have escaped detection by any of the high resolution techniques currently used for the examination of solid structures, in contrast to the readiness with which such defects are detected in quenched metals. Small spherical clusters would be very difficult to detect by methods which depend upon diffraction contrast, but etching effects are sometimes cited as evidence for their presence (see, e.g., Doherty and Davis, 1959).

The structure of a crystal grown under carefully controlled conditions will generally be much more perfect than that of a crystal in a randomly frozen aggregate. Dendritic growth may be prevalent either in pure metals or in alloys, and may be detected either by the relief pattern left on the top of a solidified ingot because of shrinkage, or by pouring out the liquid from a partly solidified ingot. In alloys, it may be seen by etching a microsection as the composition of the solid changes continuously during solidification.

The Mullins-Sekerka stability analysis described in Section 54 indicates that if the further growth of a solid nucleus is controlled by diffusion of either solute or heat, even a smoothly growing spherical crystal will develop protuberances when its diameter exceeds about  $20r_c$ , where  $r_c$  is the radius of a critical nucleus, p. 426, and this applies a fortiori to polyhedral nuclei because of the "diffusion point effect". Such instabilities may then grow into isolated dendrites or cellular or dendritic arrays. As already described in Section 54, most attempts to develop viable analytical theories of dendritic growth have concentrated on models of a freely growing isolated dendrite. These descriptions generally deal only with the geometry and growth rate of the tip of the primary dendrite and do not include directly the sideways formation of the secondary arms. However, recent work by Hunt (1991) suggests that a better procedure may be to model the growth of arrays, and then to treat an isolated dendrite as a component of an array of large spacing.

Single dendrite models are usually used on the assumption, implicit or explicit, that steady-state growth has been attained, so that any point defined with respect to the tip of a growing thermal dendrite by fixed coordinates in a system moving with velocity  $\Upsilon$  (e.g. with the origin in the moving interface) has a temperature, independent of time, which satisfies the differential equation for steady-state heat flow,

$$D_T^1 \nabla^2 T + \Upsilon (dT/dx_1) = 0 \quad (66.7)$$

equivalent to a three-dimensional version of eqn. (54.45) for material flow. None of the theories of this type lead to an equation which specifies the growth rate  $\Upsilon$  for given bath supercooling; instead, they all give a value for the product of the growth rate and the tip radius  $r$  (i.e. for the Péclet number) and some other principle has to be invoked to fix  $\Upsilon$  and  $r$  separately. For many years, Zener's maximum velocity hypothesis (see pp. 495 and 500) was thought to govern selection of the actual tip radius and velocity, but experimental results, especially those of Huang and Glicksman (1981), show beyond any possibility of error that the observed dendrites do not have the maximum possible velocity. A major advance in recent years has been the development of a

better understanding of how the system fixes on one particular tip radius and associated velocity.

The pioneering work in this field was due to Papapetrou (1935) who suggested that an isothermal dendrite having the shape of a paraboloid of revolution could preserve its shape and grow linearly under diffusion control. A proof of this for both the axially symmetric paraboloid and for a parabolic cylinder (an infinite parabolic plate) was given by Ivantsov (1947), and Horvay and Cahn (1961) generalized his solution and proved that for any dendrite with a longitudinal section in the form of a parabola and an elliptical cross-section, there exists a steady-state solution of the solute (or energy) diffusion problem, provided that the composition (or temperature) is constant over the surface of the dendrite. Such a steady-state growth model makes no provision for the formation of secondary and tertiary arms, but it has been frequently assumed to give an approximate description of the growth of the main stem of the dendrite. The discussion given on pp. 498–503 of steady-state dendritic growth applies to solidification of pure metals if the Péclet number for diffusional growth is replaced by the thermal Péclet number  $\bar{p}_{Th}^1$ , which is defined as

$$\bar{p}_{Th}^1 = \Upsilon r / 2D_{Th}^1 \quad (66.8)$$

where  $r$  is the radius of curvature of the interface (or  $r = (r_1 r_2) / \{r_1 + r_2\}$ ) if there are two different principal radii of curvature). A corresponding Péclet number,  $\bar{p}_{Th}^s$ , may be defined for the solid phase.

The solutions of Ivantsov and Horvay and Cahn are exact for the assumed boundary condition of an isothermal interface, and although this assumption is invalid, the modifications to the shape introduced by inclusion of the Gibbs–Thomson effect are very small except for very small tip radii, i.e. at high undercooling, as experienced, for example, in rapid solidification. Because the curvature undercooling is normally much smaller than the diffusion undercooling, Trivedi and Kurz (1994b) conclude that the capillary term may readily be omitted from the calculation of thermal or material diffusion fields, although, paradoxically, the Gibbs–Thomson effect is nevertheless significant in fixing the actual tip radius and velocity. This is because the Gibbs–Thomson effect and the interfacial kinetic resistance to growth both tend to stabilize a smooth interface, whilst temperature (or composition) gradients have a destabilizing effect.

According to the Ivantsov model, the dendrite tip velocity is inversely proportional to the tip radius, so there is no maximum velocity but the velocity tends to infinity as the tip radius decreases towards zero. This is inherent in a pure diffusion description, and is avoided in the so-called “modified Ivantsov” model developed by Glicksman and Schaefer (1967, 1968). It is assumed that Ivantsov’s solution remains valid for the difference in temperature between the liquid bath and the tip of the dendrite, but the tip temperature incorporates the additional capillarity supercooling appropriate to the tip curvature. The Gibbs–Thomson effect is introduced only at the dendrite tip, and the velocity versus tip radius relation (Fig. 14.10) is almost identical with that of the Ivantsov solution except for very small tip radii where, as  $r$  is reduced, the velocity reaches a maximum and then decreases rapidly.

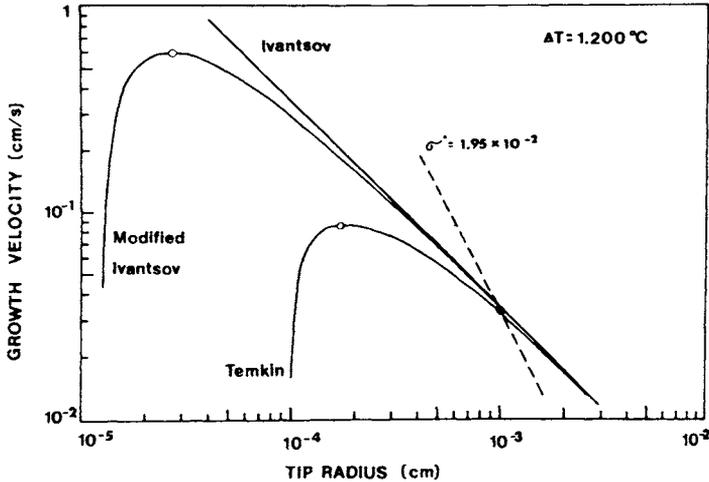


FIG. 14.10. Tip velocity versus tip radius (after Glicksman and Schaefer, 1967).

A convenient, dimensionless measure of the supercooling needed to remove the latent heat from the interface region is the quantity

$$U^* = C_p^l(\Delta T^*)/\Delta h^sl \quad (66.9)$$

where  $\Delta T^*$  is given by eqn. (66.5). The Ivantsov-Horvay-Cahn solutions can then be written in the form of eqn. (54.58) as

$$U = f(\bar{p}_{th}^1) \quad (66.10)$$

where the extreme values of the function are those given in eqns. (54.56) and (54.57). As already mentioned [see eqn. (54.59)], the theory also predicts an approximate power law relation between the Péclet number and the liquid supercooling, with the exponent  $z$  in eqn. (54.56) varying from 1.2 to 2.0 as the dendrite changes from a paraboloid of revolution to a parabolic plate.

More elaborate theories of dendritic growth originated with Temkin (1960) and Bolling and Tiller (1961), and were further developed by Kotler and Tarshis (1968, 1969) and by Trivedi (1970). All these treatments discard the assumption that the interface is isothermal (or isoconcentrate) but they retain the approximation that the shape of the dendrite is paraboloid. Although this is now generally acknowledged to be incorrect, more rigorous calculations (mentioned below) show that the actual shape deviates very little from a paraboloidal shape. Trivedi applied the curvature undercooling to all parts of the interface, so that the temperature of the interface varies on moving back from the tip. His treatment removed some approximations in the work of Temkin and in that of Kotler and Tarshis, and his solution to the heat flow problem is exact within the assumed boundary conditions. This solution is given by the thermal equivalent to eqn. (54.66) for a plate-shaped dendrite and the modifications listed on p. 503 give the corresponding solution for a paraboloid of revolution.

Nash and Glicksman (1974ab) emphasized that growth of a non-isothermal dendrite, even with the assumption that a steady-state condition has been obtained, is essentially a free boundary problem. The dendrite shape must be found as part of the solution, and the assumption of a particular shape such as a paraboloid of revolution is not consistent with the required boundary conditions; for example, energy is not conserved at the interface. Such free boundary problems are very difficult to solve but, by using numerical methods and a linearized equation, Nash and Glicksman (1974ab) were able to obtain an approximation to a fully self-consistent solution for the particle shape and the associated diffusion field; they considered an axially symmetric dendrite with growth limited by energy diffusion and by the Gibbs–Thomson effect (with the surface free energy assumed to be isotropic), and they showed that the interface shape deviates only slightly from a paraboloid of revolution but must satisfy a non-linear integro-differential equation. If the dendrite and the liquid are assumed to have the same thermal properties, this equation contains only two dimensionless parameters,  $\Delta U^-$  and

$$\lambda = \Upsilon \sigma^{sl} C_p^l v^l / 2D_{Th}^l \Delta S^{sl} \Delta h^{sl} = \frac{1}{2} \bar{P}_{Th}^l U_r \quad (66.11)$$

For a specified undercooling, Nash and Glicksman found a family of solutions each with a particular value of  $\lambda$  (i.e. of growth rate) together with a compatible shape. This degree of freedom was expected and can be compared with the family of growth rate–tip radius solutions at fixed shape, but later work has thrown considerable doubt on this result. The theory deals only with the long time solution of a time-dependent problem and some other principle (e.g. that only dendrites of maximum growth rate ultimately survive) is needed to give a unique prediction.

Further mathematical treatment requires numerical solution of the linearized Nash–Glicksman equation but, in view of later developments, the rather complex procedures need not be described here. Several other authors (see, e.g., Meiron, 1986; Barbieri *et al.*, 1987; Kessler and Levine, 1986; Saito *et al.*, 1988) have solved the free boundary self-consistent diffusion problem, and all have concluded that the actual shape is quite close to the paraboloidal shape. According to these calculations, however, a steady-state shape change cannot be found unless the surface free energy is anisotropic, thus nullifying the basic assumption of all the analytic theories described above. This unexpected result has been called the “microsolvability condition”; it is discussed below after consideration of the stability of the interface against fluctuations.

The closeness of the free boundary results to the Ivantsov shape implies that some insight might be gained by examining and comparing the predictions of Nash and Glicksman with those of the modified Ivantsov and Temkin-Trivedi theories. This may be done as there is a theoretical limit,  $\lambda_{\max}$ , to the possible values of  $\lambda$  at any supercooling, and this leads to a unique relation between maximum growth rate and supercooling which may be compared with the predictions of the Trivedi theory, and with experimental results. For small supercooling,  $U^- \leq 0.1$ , there is a power law relation between maximum growth rate and supercooling, and the Nash-Glicksman theory predicts the relation

$$\lambda_{\max} = 0.064(U^-)^{2.65} \quad (66.12)$$

The same value is predicted by the Trivedi theory, but the growth rates given by eqn. (66.11) are a factor of  $\sim 2.5$  times slower, being approximately described by

$$\lambda_{\max} = 0.25(U^-)^{2.65} \quad (66.13)$$

In the “modified Ivantsov” theory developed by Glicksman and Schaefer (1967, 1968), the growth rate is assumed to be given by the isothermal solution, but with  $U^-$  replaced by  $U_{\text{Th}}$ . This gives a slightly smaller value of 2.21 for  $z'$ , and a growth equation

$$\lambda_{\max} = 0.041(U^-)^{2.21} \quad (66.14)$$

At fixed dimensionless supercooling of the liquid, or at fixed tip radius, the modified Ivantsov approach gives the highest growth rates and the Trivedi theory the lowest growth rates (Fig. 14.10). The Nash-Glicksman and Trivedi curves are almost parallel at small  $U^-$ , as is implied by the above equations, but at higher supercooling, the  $\log(\lambda_{\max}) - \log U^-$  relation becomes non-linear and approaches the modified Ivantsov curve. The radius of curvature at the tip of the dendrite in the Nash-Glicksman theory is approximately one-half of the isothermal tip radius for the same growth velocity, and is in good agreement with the radius predicted by the modified Ivantsov theory for values of  $U^-$  up to 0.6. In general, the modified Ivantsov theory should become a better approximation at high supercooling.

Following the failure of the maximum velocity hypothesis to explain experimental results, attention turned to considerations of interface stability against local fluctuations in temperature or, in the case of alloys, of composition. The pioneering work of Mullins and Sekerka (1963, 1964) in this field has already been outlined; later applications to dendritic growth were made by Langer and Müller-Krumbhaar (1977, 1978abc). They considered a small perturbation of the parabolic shape of an Ivantsov dendrite, and showed that the growth would be unstable against side projections for all tip radii smaller than that corresponding to the maximum velocity. Another prediction was that tip radii larger than some critical value would also be unstable. Above this critical radius, the stabilizing effect of the interfacial free energy is no longer able to prevent the dendrite from splitting into two, and this is generally called the marginal stability condition. Thus, according to this stability analysis, the actual tip radius must always be between the radius of maximum velocity (“absolute stability”) and that of marginal stability. Their analysis was for tip growth only, but they suggested that the tendency to instability behind the tip which eventually leads to the formation of side-arms would tend to displace the first unstable regions at low tip radius towards the marginal stability radius, and thus steady-state growth at the tip will correspond to the marginal stability condition.

Available experimental results seem to be in very good agreement with this marginal stability criterion, in contrast to the total failure of the maximum velocity hypothesis. Despite the argument above, however, there is no convincing reason why the marginally stable state should be selected in preference to any other of the steady-state solutions. Indeed, later analyses suggest that with a finite surface energy, there may not be a range of tip radii which correspond to steady-state, shape-preserving growth, but a unique radius

and growth velocity satisfying this requirement. A selection criterion for the marginally stable condition may be formulated in terms of a “tip selection parameter”,  $\sigma^*$ , originally introduced by Langer and Müller-Krumbhaar. A general analysis developed by Lipton *et al.* (1987) considers a balance between the thermal and/or solute gradients in the liquid, which tend to destabilize an interface, and the surface free energy, which has a stabilizing effect. In an alloy, when both thermal and solute gradients have to be considered, the rather complex general condition is given as

$$m(dc/dx)_t \zeta_c - (dT/dx)_c = \Gamma'/\sigma^* r^2 \quad (66.15)$$

where  $m$  (in an alloy) is the liquidus slope,  $(dc/dx)_t$  is the concentration gradient at the dendrite tip,  $\zeta_c$  is the function of the solute Péclet number,  $(dT/dx)_c$  is the effective temperature gradient, and is given in terms of the thermal conductivities in liquid and solid,  $K^l$  and  $K^s$ , the actual temperature gradients,  $(dT/dx)^l$  and  $(dT/dx)^s$ , and functions of the thermal Péclet numbers in liquid and solid,  $\zeta^l$  and  $\zeta^s$ , by  $(dT/dx)_c = [K^l(dT/dx)^l \zeta^l + K^s(dT/dx)^s \zeta^s]/(K^l + K^s)$ .

For thermal dendrites in a pure undercooled substance, with an isothermal (Ivantsov) interface, this criterion simplifies to

$$\Upsilon r^2 = [2D_T \Gamma' C_p^l / \sigma^* \Delta h] [\beta/\zeta] \quad (66.16)$$

where  $\beta$  is unity if the thermal conductivities of solid and liquid are equal,  $\zeta$  is effectively unity for low velocities and  $\sigma^*$  is a constant. For growth into a supercooled liquid, the effective thermal gradient  $(dT/dx)_c$  is negative, so that the two gradients (in the case of an alloy) on the left-hand side of eqn. (66.15) are additive and together represent the total destabilizing effect, which is balanced by the capillary effect on the right-hand side of the equation. In the case of directional solidification, the effective thermal gradient is positive, and thus stabilizing. The sign of the difference in the two gradients then gives an approximate condition for the formation of a cellular microstructure (see below). According to Trivedi and Kurz, the dendrite tip selection parameter,  $\sigma^*$ , when properly defined by eqn. (66.15), has a constant value in a particular system and is not, for example, a function of velocity or of composition. Langer and Müller-Krumbhaar (1978) predicted its numerical value to be about  $1/4\pi^2 = 0.025$ . Trivedi and Kurz (1994b) report various experimental determinations (see below), ranging in different materials from 0.01 to 0.08.

A significant consequence of eqn. (66.16) with constant  $\sigma^*$  is that if  $(\beta/\zeta)$  can be equated to unity, i.e. excluding high velocity conditions,  $\Upsilon r^2$  is predicted to have a constant value independent of the supercooling. A similar result holds in directionally solidified dendritic growth of alloys, where for moderate velocities the different constraints imply the result that  $\Upsilon r^2$  is constant, independent of the imposed velocity  $\Upsilon$ .

Equation (66.15) can be rearranged to give an equation for the tip radius which, if  $(\beta/\zeta)$  is unity, may be written in terms of the characteristic lengths  $l_T$  and  $l^B$  as

$$r = (2l_T l^B / \sigma^*)^{1/2} \quad (66.17)$$

so that, as Trivedi and Kurz point out, the radius is proportional to the geometric mean of  $l_T$  and  $l^B$  as already noted in the scaling analysis.

As mentioned on p. 659 other authors who have tackled the self-consistent free boundary problem have found that, with an isotropic surface free energy, however small, there are no true steady-state solutions. When the surface energy is anisotropic, however, a discrete set of shape-preserving solutions is obtained, but only one of these is stable against tip splitting (Kessler and Levine, 1986; Bensimon *et al.*, 1987). The condition for the stable solution is very similar to that given by eqn. (66.15) except that  $\sigma^*$  is now not a constant but varies as

$$\sigma^* = \sigma_0 e^{7/4} \quad (66.18)$$

where  $e$  is a parameter specifying the variation of surface free energy with orientation. This microsolubility approach is stated to be exact for two-dimensional dendrites, but is incomplete for three dimensions. From eqn. (66.17), it follows that as the anisotropy is reduced,  $\sigma^*$  tends to zero so from eqn. (66.15) there are no steady-state solutions with isotropic surface free energy.

The theory has thus now reached a state in which the maximum velocity and the marginal stability conditions are no longer viable, and if there are steady-state growth conditions, they may represent unique conditions and shapes of the dendrite. Further light is shed on this by the recent work of Hunt and Lu (Hunt, 1990, 1991, 1997; Hunt and Lu, 1996; Lu and Hunt, 1992; Wan *et al.*, 1997).

Hunt considers the time-dependent growth of arrays, initially intended to be cellular arrays in alloys, but he discovered that the model could also be applied to dendritic arrays in pure metals. Because of the complexities of the analytical models, he used a numerical model into which he attempted to introduce all the relevant physics. The interface velocity varied over almost 12 orders of magnitude, and the undercooling over two orders of magnitude. For pure metals, the model was used for steady-state growth at very large spacings and, in agreement with the analytical models, he found the dendrites to have nearly paraboloidal shapes. However, he also found a considerable difference between dendrites with finite surface energy and those with zero surface energy. With zero surface free energy, steady-state solutions at a given liquid supercooling were found for any tip radius, but with a finite surface energy there was a unique solution at a particular radius. The experimental radius was in good agreement with the predictions of the marginal stability theory, but clearly the physical reason for the choice of radius is different. The marginal stability theory implies that one steady-state solution is being selected from a continuous range of possible solutions, whereas the Hunt model predicts that this is the only possible steady state. The experimental agreement often found with the marginal stability condition suggests that the latter might be regarded as a method of approximating the conditions under which steady-state growth is possible, rather than as a method for selecting the ultimate winner from a range of competing steady states. Because the unique steady state was obtained with isotropic surface free energy, Hunt's results clearly also contradict those aspects of the microsolubility theories which require anisotropic energies for steady-state growth. Hunt points out, however, that in his numerical model, steady-state growth has not been shown for infinitely long dendrites, but rather only for the tip region. Mathematical solutions for complete steady-state growth

may indeed be impossible, as real dendrites do not have this property but instead develop side-arms.

Direct measurements of the velocity of freely growing dendrites have been reported for only a few pure materials; early results on metals such as tin, phosphorus and nickel have been followed by more significant measurements on transparent materials such as succinonitrile, in which simultaneous information about tip geometry can be obtained. A prediction from eqn. (66.16) is that the product  $\Upsilon r^2$  is constant for all supercooling in which  $\zeta$  remains effectively unity. Huang and Glicksman (1981) made detailed measurements of tip radius and velocity on dendrites of succinonitrile and found that  $\Upsilon r^2$  was effectively constant in a range of supercooling from 1 K to 5 K. A deviation from this result at small undercooling of less than 1 K is attributed to the effects of convection. From these results, Huang and Glicksman were also able to deduce an experimental value of  $\sigma^*$  as 0.0195 and hence to calculate the variation of velocity and tip radius with undercooling predicted by the modified Ivantsov and Trivedi theories respectively over the range  $\Delta T^- = 1 - 5$  K. As already noted, the two theoretical curves are substantially in agreement over this range (see Fig. 14.11) and the experimental points deviate only slightly from the theoretical curves. This is another demonstration that the steady-state solution corresponds to the marginal stability condition, even if marginal stability is not actually involved.

Experiments on several other transparent organic materials give similar values for  $\sigma^*$  and hence do not support the solvability model as they would be expected to have different anisotropies of surface free energy, giving varying  $\sigma^*$  values according to eqn. (66.18). Some results which do not fit the theory have been explained in terms of the possible influence of sluggish interface kinetics caused by the polarity of complex molecules such as those in pivalic acid.

Many investigations have been made of the crystallography of dendritic growth. Weinberg and Chalmers (1951, 1952) and Graf (1951) independently pointed out that the preferred  $\langle 100 \rangle$  growth direction of f.c.c. crystals is the axis of a pyramid of the four  $\{111\}$  planes. The dendrite axes in other crystals are also always found to be symmetrical with the closest-packed planes which can meet at a point. These axes and the corresponding planes are listed in Table XI; note there is only a single set of  $\{0001\}$  planes in the h.c.p. structure so these cannot determine the dendrite axis.

In tin only two  $\{001\}$  planes can meet at a point and the observed dendrite axis is the line bisecting the wedge thus formed. A pyramid may be made from the  $\{100\}$  planes and the next densely packed  $\{112\}$  planes and the axis of this pyramid has been observed as a secondary dendritic axis in tin. Two of the  $\{11\bar{2}0\}$  planes of an h.c.p. structure also meet to form a wedge which may have an acute or an obtuse angle, and the corresponding bisecting line is a  $\langle 11\bar{2}0 \rangle$  or a  $\langle 10\bar{1}0 \rangle$  direction respectively.

Frank (1958) pointed out that the Weinberg and Chalmers' rule is ambiguous for b.c.c. crystals where  $\{100\}$  planes make trigonal pyramids in  $\langle 111 \rangle$  directions as well as fourfold pyramids in  $\langle 100 \rangle$  directions. He suggested that the dendrite axes correspond to the extremities of the equilibrium body according to the Wulff construction (p. 155). At low temperatures, these will correspond to cusps in the polar surface free energy plots (Fig. 5.11); at higher temperatures they will correspond to rounded minima.

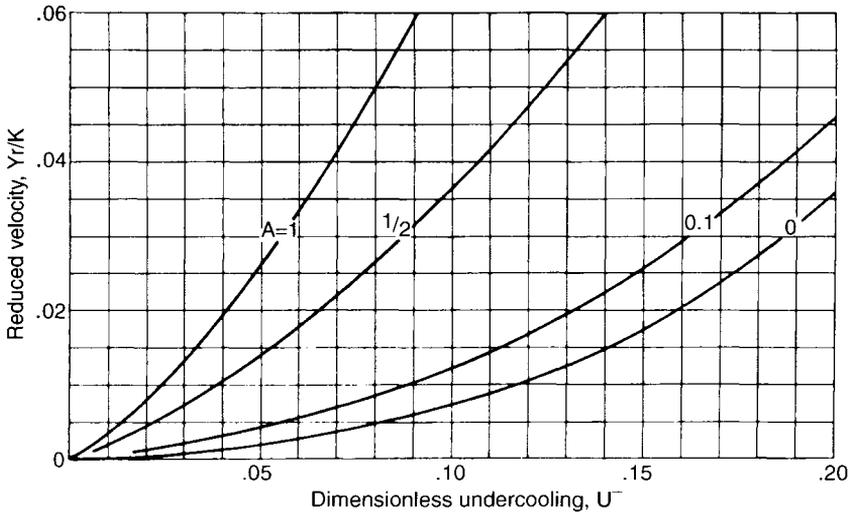


FIG. 14.11. Growth rates versus dimensionless undercooling.

TABLE XI.

Structure	Dendrite axis	Close-packed planes
f.c.c.	$\langle 100 \rangle$	$\{111\}$
b.c.c.	$\langle 100 \rangle$	$\{110\}$
h.c.p.	$\langle 10 \bar{1}0 \rangle$	$\{11\bar{2}0\}$
b.c. tetr. } (tin) }	$\langle 110 \rangle$	$\{100\}$
	$\langle 001 \rangle$	$\{100\} + \{112\}$

Rosenberg and Tiller (1957) used decanting experiments to verify the transition from normal growth to dendritic growth. A planar interface of zone-refined lead with a single system of  $\{111\}$  platelets appeared to develop other platelets when dendrites grew out. Photographs of the dendritic needle usually showed four sets of platelets extending to the tip although occasionally  $\{100\}$  platelets were also visible. However, as in the case of planar interfaces, these stepped or platelet structures may be artefacts resulting from retained liquid on the decanted surface.

## 67. GROWTH FROM A LIQUID ALLOY

Very small quantities of a second component in the melt of a nominally pure metal may have important effects on the growth of the solid. Some of the phenomena to be described in this section may thus be relevant not only to deliberate alloys but also to metals which are not of the highest purity. Consider first the simple case where the equilibrium structure

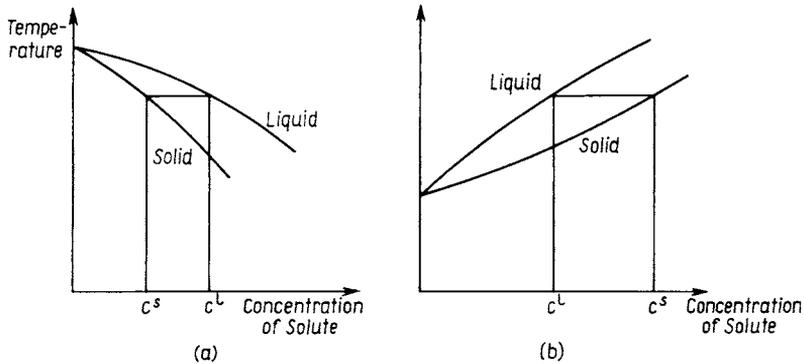


FIG. 14.12. The distribution coefficient.

at temperatures immediately below the solidus is a homogeneous solid solution. Except at extremely slow solidification rates, a casting will not be homogeneous but will have a “cored” structure as a result of the difference in equilibrium compositions of the solid and liquid in the two-phase region. The important parameter defining the segregation is the distribution coefficient  $i$ , which is the ratio of the concentration of solute in the equilibrium solid phase to that in the equilibrium liquid phase. In general,  $i$  will vary with the composition of the alloy but, in sufficiently dilute alloys, the liquidus and solidus may be approximated by straight lines and  $i$  may be treated as constant. If the freezing point of the pure solvent is lowered by addition of solute,  $i$  is less than unity; if it is raised,  $i$  is greater than unity (Fig. 14.12).

In principle, segregation may be avoided if freezing is slow enough to maintain thermodynamic equilibrium at all stages, but this is scarcely possible in practice. To maintain equilibrium at all times, freezing must begin when the temperature crosses the liquidus, and the first solid deposited will have a composition  $i$  times the overall composition,  $c_0$  solute atoms per unit volume, of the alloy. During the subsequent growth of the solid, its composition must be kept uniform by diffusion as layers of increasing or decreasing solute content (depending on whether  $i$  is less than or greater than unity) are deposited on it from the liquid. The composition of the solid thus moves continuously along the solidus line,  $T^{sl}(c)$ , whilst that of the liquid moves along the liquidus line,  $T^{ls}(c)$ , until the solidus temperature of the original composition is reached and the last drops of liquid, having composition  $c_0/i$ , solidify to give a uniform solid of the original composition. Diffusion over macroscopic distances is required to keep the concentration of solute uniform in the solid, and this is rarely sufficiently rapid to keep pace with the rate of solidification.

In practice, it is usually a better approximation to use the opposite assumption, namely that diffusion in the solid is negligible during the period of solidification. The amount of segregation then depends on the growth rate which determines the extent of mixing in the liquid. Solute distribution in the liquid is influenced both by diffusion and by stirring and natural convection. The two extremes of perfect mixing in the liquid and of transport only

by liquid diffusion may both be treated readily; under most conditions of growth, there is probably partial mixing somewhere between these extremes.

There is an important distinction to be made between alloy solidification experiments and the corresponding experiments on precipitation of one phase from another in the solid state. As already noted, growth during solidification is rapid and in many circumstances may be governed by the rate of heat removal. This means that isothermal solidification cannot readily be achieved, and controlled growth experiments usually involve directional solidification, in which the free variables are the temperature of the interface and the compositions of the liquid and solid phases at the interface. In alloys, these compositions are typically diffusion-controlled, but it is often convenient to express the boundary conditions simply in terms of the interface temperature as it is again a basic assumption of almost all growth theories that the two phases are in quasi-equilibrium with each other at the freezing rates typical of controlled freezing. So long as the interface is planar, the temperature variation is unidirectional, and the solutions of Section 54 for diffusion-controlled growth are readily adaptable to the non-isothermal steady state. A full theory of dendritic type growth, where in principle both solute and temperature fields become three-dimensional, is much more complex and approximations have to be used.

Consider the freezing of a liquid alloy in which the initial concentration of solute atoms per unit volume is  $c_0$  and assume that the liquid concentration is uniform at all times. Let the concentration in the solid be  $c^s$  when a fraction  $\zeta$  of the liquid has solidified. As this solid is in equilibrium with the liquid, the concentration in the latter is  $c^l = c^s/i$  and the total amount of solute remaining in the liquid is  $V(1 - \zeta)c^s/i$ , where  $V$  is the volume of the whole assembly. Now let a further small volume  $Vd\zeta$  solidify. The concentration of solute in this volume is given by

$$c^s = -(1/V)[d\{V(1 - \zeta)c^s/i\}/d\zeta]$$

The first solid formed has composition  $c^s = ic_0$ , so this equation may be integrated between limits  $c^s = ic_0$ ,  $\zeta = 0$  and  $c^s = c^s$ ,  $\zeta = \zeta$  to give

$$c^s = ic_0(1 - \zeta)^{i-1} \quad (67.1)$$

Although apparently first derived by Gulliver (1913), this is usually known as the Scheil equation (Scheil, 1942).

The derivation of eqn. (67.1) neglects any change of volume on freezing and also assumes perfect liquid mixing and a constant value of  $i$ . The result is independent of the shapes of the solid and liquid regions during freezing, but it is most readily applied to the unidirectional freezing of a column of liquid. The parameter  $\zeta$  is then proportional to the length of the solidified portion and the equation gives the solute distribution over most of the final ingot. However, the equation is clearly only an approximation because it predicts  $c^s = \infty$  at  $\zeta = 1$  which is clearly impossible. In an actual assembly,  $i$  must either change with composition to allow the liquidus and solidus lines to meet again, or a eutectic or peritectic reaction must intervene. With reasonably dilute solutions, the equation is nevertheless a useful approximation until about 80% of the liquid has solidified.

The liquid concentration remains uniform only for very slow rates of growth. As solid is deposited, extra solute is effectively rejected to the liquid (if  $i < 1$ ) or absorbed from it (if  $i > 1$ ) and if this cannot diffuse sufficiently rapidly to maintain a uniform composition, there will be a layer of liquid adjacent to the interface which is enriched or depleted in solute. The solute concentration in this layer, rather than that in the main body of the liquid, then determines the concentration in the solid being deposited at any stage. As the concentration of solute in the liquid changes, so does that in the deposited solid, and a steady state is reached when the rate at which extra solute is rejected to (or absorbed from) the liquid just equals the rate at which it can be dispersed or supplied by diffusive and convective mixing. The ratio of the concentration of solute in the deposited solid to that in the liquid remote from the interface can be used to define an effective distribution coefficient which initially has the equilibrium value  $i$ , but approaches asymptotically a limiting value  $i'$  as the steady state is reached. In the extreme case when transport in the whole liquid is by diffusion only, the limiting value is  $i' = 1$  and the effective distribution coefficient approaches this value if the melt is large enough. The solute distribution under these conditions was analysed by Tiller *et al.* (1953). Figure 14.13 shows qualitatively the distribution of solute in the liquid (for  $i < 1$ ) after attainment of the steady state.

Consider the origin of the  $x$  axis to be in the solid-liquid interface and consider a volume element bounded by planes of unit area normal to this axis at distances  $x$ ,  $x + dx$  from the interface. The net flow of solute by diffusion into this element in time  $\delta t$  is  $D^l(\partial^2 c/\partial x^2)\delta t$  and the net flow of solute out of the element due to freezing (i.e. moving the liquid distribution relative to the interface) is  $\Upsilon(\partial c/\partial x)\delta t$ . Thus in the steady state

$$D^l(\partial^2 c/\partial x^2) + \Upsilon(\partial c/\partial x) = 0 \quad (67.2)$$

which has the solution

$$c = c_0 + c_1 \exp(-\Upsilon x/D^l)$$

where the boundary condition  $c = c_0$  at  $\zeta = 0$  has been used. To determine  $c_1$ , note that when the steady state is reached the solid composition at the interface must be  $c_0$  and that

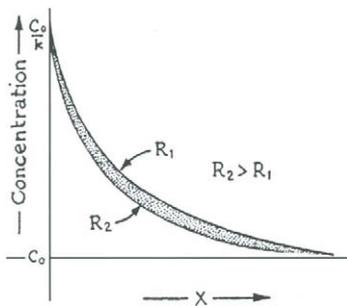


FIG. 14.13. Distribution of solute in liquid after attainment of the steady state (after Tiller *et al.*, 1953).

of the liquid must be  $c_0/i$ . Substituting for  $c_1$  gives

$$c = c_0[1 + (i^{-1} - 1) \exp(-\Upsilon x/D)] \quad (67.3)$$

for the steady-state solute distribution in the liquid. Equation (67.3) may also be written in the alternative form

$$c = (D/\Upsilon)(dc/dx)_l\{1 - \exp(-\Upsilon x/D)\} + c^l \quad (67.4)$$

where  $(dc/dx)_l$  is the composition gradient in the liquid at  $x=0$ , i.e. immediately adjacent to the interface, and  $c_0 = ic^l = c^l + (D^l/\Upsilon)(dc/dx)_l$ .

Now consider the heat flux which is coupled to the solute flux. If conduction of heat in both liquid and solid is taken into account, there are two differential equations analogous to eqn. (67.2)

$$D_T^l(\partial^2 T/\partial x^2) + (\Upsilon\partial T/\partial x) = 0 \quad (x > 0) \quad (67.5a)$$

$$D_T^s(\partial^2 T/\partial x^2) + (\Upsilon\partial T/\partial x) = 0 \quad (x < 0) \quad (67.5b)$$

giving the temperature distribution in the liquid and the solid. There is a boundary condition that at the interface  $T = T^{ls}(c^l) = T^{ls} + mc^l$ , where  $T^{ls}(c^l)$  is the liquidus temperature for composition  $c^l$  and  $T^{ls}$  is the freezing temperature for the pure solvent. Equations (67.5) now lead to temperature distributions

$$T = (D_T^l/T)(dT/dx)_l\{1 - \exp(-\Upsilon x/D_T^l)\} + T^{ls}(c^l), \quad (x > 0) \quad (67.6a)$$

and

$$T = (D_T^s/T)(dT/dx)_s\{1 - \exp(-\Upsilon x/D_T^s)\} + T^{ls}(c^l), \quad (x < 0) \quad (67.6b)$$

which with the obvious notation  $(dT/dx)_l$  and  $(dT/dx)_s$  for the two temperature gradients at the interface are the thermal equivalents of eqn. (67.4). There are now two equations instead of one because the thermal diffusivity in the solid, unlike the material diffusivity, is not negligible. Moreover, as  $D_T^l/D$  is typically of the order of  $10^4$ , the temperature variation is almost linear over the whole range of  $x$  in which  $c$  varies. Figure 14.14 shows schematically the variation of composition with distance in the vicinity of the interface for two different interface positions. Steady-state diffusion-controlled flow is strictly only possible when the experimental arrangements allow both eqns. (67.4) and (67.6) to be satisfied. Note also that conservation of solute [see eqn. (54.3)] requires that

$$\Upsilon = D(dc/dx)_l/(c_0 - c^l) = (v^s/\Delta h^{sl})[K_T^{sl}(dT/dx)_{-1} - K_T^l(dT/dx)_l] \quad (67.7)$$

Expressions may also be derived for the solute distribution in liquid and solid during the period in which the steady state is established. Equation (67.2) has now to be replaced by the time-dependent equation

$$D^l(\partial^2 c/\partial x^2) + \Upsilon(\partial c/\partial x) = \partial c/\partial t \quad (67.8)$$

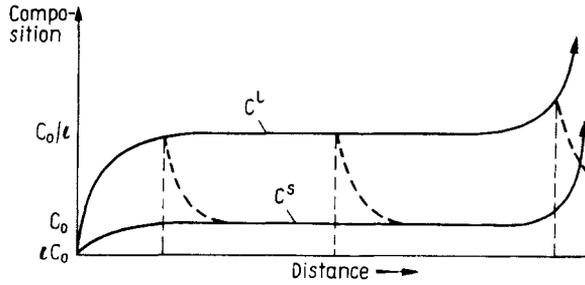


FIG. 14.14. The distribution of solute in liquid and solid when liquid transport is by diffusion only (after Tiller *et al.*, 1953).

with initial conditions  $c = c_0$  at  $t = 0$  and boundary conditions  $c = c_0$  at  $x = \infty$ . The complete solution is rather complex (Smith *et al.*, 1955) but to a good approximation for small values of  $i$  the solute concentrations at the interface are given by

$$c^s = ic^l = c_0 \{1 - (1 - i) \exp(-iYx/D)\} \quad (67.9)$$

The solute content of the solid increases gradually from its initial value of  $ic_0$  to its steady-state value of  $c_0$  and the effective distribution coefficient  $i' = c^s/c_0$  increases from  $i$  to unity in a distance which depends on the value of  $D/iY$ . Thus, for small  $i$ ,  $i'$  has risen to  $\sim 0.63$  and  $\sim 0.86$  in distances of  $D/iY$  and  $2D/iY$  respectively. There is also of course a terminal transient when the boundary condition at  $x = \infty$  no longer applies. This is much shorter than the initial transient as its characteristic length is determined by the thickness  $D^1/Y$  of the solute boundary layer.

The final distribution of solute in the two extremes of complete mixing in the liquid and transport in the liquid by diffusion only is shown qualitatively in Fig. 14.15. The lengths of the initial and final portions of curve (b) are arbitrary, but there is a considerable difference between the two curves. Most actual freezing conditions correspond to incomplete convective mixing of the liquid and so lead to a distribution between these two curves. Burton *et al.* (1953) developed a model for the case of partial mixing in which solute redistribution in the liquid is entirely by diffusion within a boundary layer of thickness  $\delta^B$ , whilst outside this layer, the composition is maintained uniform by convection. Their analysis shows that after an initial transient during which the solid composition changes from  $ic_0$  to  $i'c_0$ , the distribution of solute along a solidified bar may be represented by a modified form of the Scheil equation (67.1)

$$c^s = ic_0(1 - \zeta)^{i'-1} \quad (67.10)$$

in which the effective distribution coefficient is given by

$$i'/i = [i + (1 - i) \exp(-Y\delta^B/D)]^{-1} \quad (67.11)$$

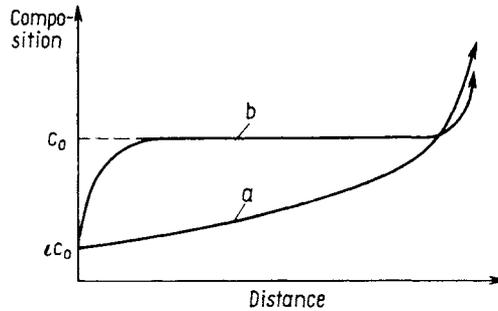


FIG. 14.15. Final distribution of solute after normal freezing for the extreme cases of complete mixing in the liquid (curve a) and transport by diffusion only (curve b).

Tests of this theory have been made by measuring the solute distribution along a bar which had been solidified unidirectionally at various growth rates. Dean *et al.* (1962) worked with a Pb–1% Sn alloy and found that the measured distribution could be fitted to eqn. (67.10) and values of  $i'$  thus deduced. The growth rate and the temperature gradient ahead of the interface were varied independently, and it was found that planar interface solidification was favoured by low growth rates and high temperature gradients. The effective distribution coefficient  $i'$  varied with growth rate according to eqn. (67.11) and was independent of temperature gradient. From these results, the boundary layer thickness was calculated as  $\delta^B \cong 0.5$  mm. Zief and Wilcox (1967) have computed other values of  $\delta^B$  from experiments on normal solidification, Czochralski growth and zone melting; in metallic systems, these range from 0.07 mm in Mg 0.0016% Fe to 0.87 mm in Pb 10% Sn. Sharp and Hellawell (1971) found values between 0.25 and 0.42 mm in Al–Cu alloys with varying compositions and temperature gradients.

It follows from eqns. (67.10) and (67.11) that the effective distribution coefficient and the solute content in the solid are both raised by increasing growth rate or boundary layer thickness. Fluctuations in growth rate produced, for example, by temperature oscillations, may thus produce fluctuations of composition in a growing crystal, a phenomenon which is sometimes known as “banding”.

Another important distribution arises from zone-melting procedures, in which a long solid charge is traversed by one or more short molten zones. As such a zone is moved along the crystal, a melting interface at any point is followed by a solidification interface so that, in contrast to normal freezing, only a small part of the assembly is liquid at any time. Zone melting is a useful technique for growing single crystals, especially when “floating” zones (without crucibles) may be used, as for many refractory metals. Effective procedures for controlling the solute distribution in any desired way have been devised; they range from impurity separation (“zone refining”) to homogenization (“zone levelling”).

Consider a solid rod of uniform composition which is traversed by a single molten zone of length  $l_z$ . At the beginning of the zoning, there is a liquid layer containing solute of concentration  $c_0$  but, as the zone moves forward a short distance, a new solid layer of composition  $ic_0$  is deposited and a corresponding amount of the original solid is melted.

The concentration of the liquid zone is thus enriched in solute (for  $i < 1$ ) and the concentration of the solid formed at the freezing interface correspondingly increases. The enrichment continues at a decreasing rate until the liquid concentration reaches  $c_0/i$ . The concentration of solute in the freezing solid is then  $c_0$ , so that a steady state has been attained, and the zoned solid has uniform composition until the zone approaches the end of the rod. The last liquid of length  $l_z$  solidifies in the manner described above for normal freezing, and the solute concentration rises again.

The solute distribution produced by zoning is similar to that predicted for normal freezing without convection; the use of a short zone may be regarded as a device to prevent mixing in the liquid region. Figure 14.16 shows the distribution of solute after a single zone pass for  $i < 1$ . The first two parts of the curve may be represented by the equation

$$c^s = c_0[1 - (1 - i) \exp(ix/l_z)] \quad (67.12)$$

The derivation of this equation, due to W.T. Read (1951), is similar to that of eqn. (67.1), and the distribution in the third region is actually given by eqn. (67.1).

The distribution of solute after several zones have been passed through the rod in the same direction is difficult to analyse mathematically, but it can readily be seen that the second zone will accumulate solute, thus leaving a longer and purer initial region, and the solute content will begin to rise again after the plateau region when the zone is within  $2l_z$  of the end of the rod. With further zone passes, both transient regions spread towards the centre and eventually join together, thus eliminating the plateau. The majority of the solute has been "pushed" to the end of the rod, which is the desired result in zone refining. An opposite effect is produced if the second zone is passed along the rod in the reverse direction, this being one of the procedures used to produce zone levelling. An early, but very detailed, account of all zoning procedures was given by Pfann (1957).

The uneven distribution of solute (or impurity) atoms produced in the solid by normal freezing has various secondary consequences. One of these is that the natural lattice parameter will vary with position in the final crystal or rod, thus producing internal stresses because of coherency constraints. These stresses may be large enough to generate dislocations, as first discussed in detail by Frank (1952) and by Tiller (1958).

The difference in composition between the liquid adjacent to and remote from the solid-liquid interface leads to the important concept of constitutional supercooling (Rutter and

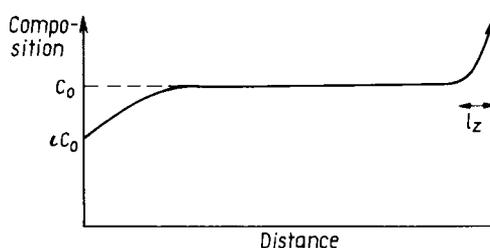


FIG. 14.16. Distribution of solute after one zone pass (after Pfann, 1952).

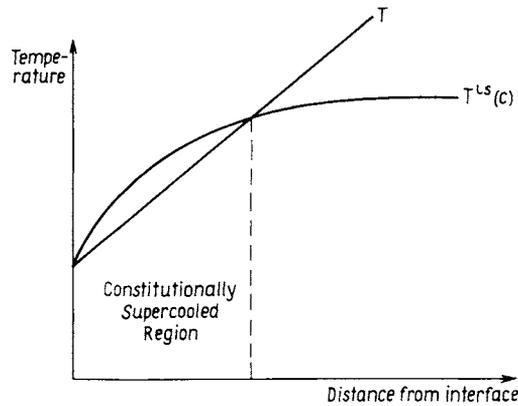


FIG. 14.17. Diagram illustrating the formation of a constitutionally supercooled layer of liquid (after Rutter and Chalmers, 1953).

Chalmers, 1953). For  $i < 1$ , the liquid adjacent to the interface has a higher solute content than the rest of the liquid (Fig. 14.12), and consequently has a lower freezing temperature than the remainder of the liquid. The equilibrium liquidus temperature thus varies with the local composition, and rises to a constant value in liquid remote from the interface. Figure 14.17 shows this liquidus (freezing) temperature,  $T^{L_S}$ , and also the actual temperature,  $T$ , as a function of distance from the interface. Although the temperature rises continuously, there may nevertheless be a region in which the actual temperature is below the temperature at which (equilibrium) freezing begins for an alloy with the local composition. This region is described as "constitutionally supercooled".

The existence of a constitutionally supercooled region has effects similar to those previously described for a temperature inversion, and a planar or smoothly curved solid-liquid interface will tend to be unstable. A protrusion formed by some fluctuation may continue to grow further ahead of the original interface because it is experiencing an increased driving force for growth; thus, under certain conditions, dendrites will grow out into the liquid. Thus one consequence of constitutional supercooling is the possibility of dendritic growth even when the liquid ahead of the dendrite is hotter than the dendrite itself. This provides one possible explanation of how dendrites can grow towards the centre of a cast ingot which is being cooled from the outside (see Section 68).

It is clear from Figure 14.15 that the extent of the constitutionally supercooled layer depends on the temperature gradient at the interface, decreasing as the gradient is increased and disappearing altogether when the gradient exceeds the initial slope  $(dT^{L_S}/dx)_{x=0}$  of the liquidus curve. In conditions such as slow directional freezing, the constitutionally cooled region only extends for a short distance ahead of the interface and any projections can only grow forward this short distance before they encounter liquid which is too hot to solidify. Rutter and Chalmers (1953) showed that under these conditions a stable liquid-solid interface consisting of a number of rounded projections can grow into the liquid. Diffusion of solute takes place in lateral directions away from the

liquid near the projecting tips (for  $i < 1$ ) into the regions between the projections. This enables the tips to grow at a higher temperature than the valleys and thus stabilizes the non-planar interface. This cellular substructure was first identified by Smialowski (1937), who correctly ascribed it to segregation effects. Before Smialowski's work, Buerger (1934) had shown a photograph of a zinc crystal with what appears now to be a cellular substructure but he believed it to be a misorientation (or lineage) structure of the type already discussed. However, the two substructures are not unrelated; when a cellular structure forms in alloys, there is a strong tendency for any low angle boundaries to be associated with the intercell boundaries.

The Smialowski structure, as it is sometimes called, usually appears as a honeycomb pattern of cells which are approximately hexagonal when observed on the growth interface; an example is shown in Fig. 14.18. These cells are separated by the intercell regions of liquid which extend a considerable distance back from the tip position. The cells are elongated in the growth direction, and the structure appears as narrow furrows on a free surface containing the growth direction. The whole crystal is thus divided into hexagonaloid "pencils" separated from each other by boundaries relatively rich (or poor if  $i > 1$ ) in solute.

Later experiments suggested that there are intermediate morphologies between the stable planar and hexagonaloid cellular interfaces. There is also a considerable difficulty in distinguishing between a cellular and a dendritic interface. The simplest definition has been that if side-arms develop, the structure is dendritic, whilst projections without side-arms are defined as cellular. A less obvious but more fundamental distinction is in the shape of the tip of the projecting region. As already emphasized, this is near parabolic in a longitudinal section of a dendrite, but is much flatter in the cellular structure.

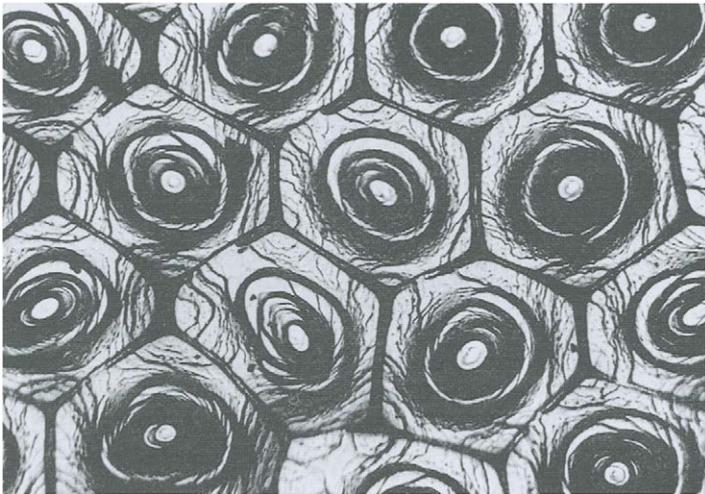


FIG. 14.18. Cellular structure on the decanted surface of a Pb-0.1% Sn crystal (original magnification  $\times 300$ ). Growth rate  $10^{-2} \text{ mm s}^{-1}$  (after Hellowell, 1962).

According to the theory of constitutional supercooling, the planar interface will remain stable as long as the thermodynamically stable state of the material immediately ahead of the interface is liquid, but it will become unstable when the equilibrium state ahead of the interface is solid. This thermodynamic criterion is only an approximation, and kinetic effects must be considered in a proper assessment of interface stability. If solute transport in the liquid is mainly by diffusion, the concentration of solute will increase exponentially with distance, as in eqn. (67.3). The distance at which the excess solute concentration has fallen to  $1/e$  of its value at the interface will roughly correspond to the diffusion layer thickness, and is thus  $l^B = D^I/\Upsilon$ , which was previously introduced as a characteristic diffusion length.

A qualitative argument shows that the wavelength of the stable corrugations on the liquid–solid interface must be comparable in magnitude to  $l_D$  as diffusion over greater distances during growth is not possible, and a finer structure would have more liquid–solid interfacial area, and hence higher energy. Reasonable values of  $D^I$  ( $10^{-9} \text{ m}^2 \text{ s}^{-1}$ ) and  $\Gamma$  ( $10^{-5} \text{ m s}^{-1}$ ) give  $l^B = 0.1 \text{ mm}$ , in agreement with the observed scale of the structure.

This general explanation of the cellular structure was first verified by Rutter and Chalmers (1953) who found no such structure in very pure metals, but a well-developed structure with as little as 0.01% impurity. Autoradiographic techniques were used in the early work to verify solute segregation to (or from) the inter-cell grooves.

According to the simple theory of constitutional supercooling, a macroscopically smooth interface should become unstable at a critical ratio of the temperature gradient,  $(dT/dx)_I$ , in the liquid at the interface to the rate of growth,  $\Upsilon$ . For dilute solution, the liquidus temperature at composition  $c$  is given by  $T^{ls}(c) = T^{ls}(0) + mc$ , where  $T^{ls}(0)$  is the freezing point of the pure metal. Substituting for  $c$  from eqn. (67.3), the variation of liquidus temperature with distance from the interface is

$$T^{ls}(c) = T^{ls}(0) + mc_0 \left[ 1 + (i^{-1} - 1) \exp(-\Upsilon x/D^I) \right] \quad (67.13)$$

The actual temperature at a distance  $x$  from the interface is

$$T = T^{ls} + mc_0/i + x(dT/dx)_I \quad (67.14)$$

It follows that  $T = T^{ls}(c)$  when

$$[1 - \exp(-\Upsilon x/D^I)] = -i(dT/dx)_I x / [mc_0(1 - i)] \quad (67.15)$$

When this equation has a real solution, other than  $x=0$ , there is a constitutionally supercooled region extending from the interface to the  $x$  value given by eqn. (67.15) which is the intersection coordinate of the two curves shown in Fig. 14.17. The two curves intersect if the tangent at the origin to  $T^{ls}$  has a higher slope than that of the actual temperature. This predicts constitutional supercooling if

$$(dT/dx)_I \Upsilon < -(mc_0/D^I)(i - 1)/i \quad (67.16)$$

Note that the critical value of  $(dT/dx)_I$  when this inequality becomes an equation is necessarily positive because  $m$  and  $(1 - i)$  must have opposite signs.

An early study of the transition from planar to cellular growth was made by Walton *et al.* (1955) who used alloys of up to 0.02% pure lead in very high purity (zone-refined) zinc. With various growth rates and temperature gradients, they found a straight line relation between  $c_0$  and  $(dT/dx)/\Upsilon$ , from the slope of which they estimated the value of the diffusion coefficient  $D^l$  of lead in liquid tin to be  $2 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ , in very good agreement with measured values at higher temperatures.

As noted by Walton *et al.* and confirmed in many later investigations, there is some ambiguity in experiments of this kind because of the succession of interface morphologies observed as the interface becomes more unstable. As the ratio of the temperature gradient to the growth rate decreases, small depressions ("nodes") or protrusions ("pox") form first on the interface, then irregular cells begin to form and are followed by elongated cells and finally by the regular cell structure of Fig. 14.18. Although the transitions between these morphologies are characterized by approximately linear plots of  $(dT/dx)_I/\Upsilon$  against  $c_0$ , the structures often coexist and the transition may depend on the orientation of the solid crystal. Nodes are small circular depressions in the solid and were first identified by Biloni *et al.* (1966), who showed that they are associated with segregation of solute, unlike the small projections which had originally been thought to be the first signs of the breakdown of the planar interface. The significance of the pox structure is uncertain; it has been considered to be an artefact arising from the examination of decanted interfaces, but Biloni *et al.* believe that small projections actually form but are not connected directly with the transition. It seems probable, however, that many of the structures described as pox by earlier workers were actually examples of node structures.

Cole and Winegard (1969), using improved techniques, examined the transition in tin-lead, tin-antimony and tin-lead alloys under conditions of upward unidirectional freezing which ensured that  $i' = 1$ . They chose the first appearance of the pox structure as the criterion for the onset of constitutional supercooling in preference to the regular cell structure used by Walton *et al.*, and they deduced values of  $4.5 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$  and  $3.0 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$  for the diffusion coefficients of lead and antimony respectively in liquid tin. The transition in the ternary alloys was also found to satisfy eqn. (67.16) but the slope of the transition line indicated an interaction between the two solutes.

In a detailed investigation of the formation of cellular structures in dilute binary solutions of antimony in zone-refined lead, Morris and Winegard (1969ab) found that the planar interface becomes unstable near defects such as the container surface or grain or striation boundaries at lower growth velocities than those required to produce instability in defect-free regions. At a grain boundary, for example, a deep groove with associated solute segregation develops at the boundary and the accompanying depletion of solute in regions adjacent to the boundary leads to the formation of slight projections on either side of the groove. When the amplitude of the projections is high enough, a second groove begins to develop outside it, and in this way the instability spreads out from the grain boundary as ripples. In contrast, nodes form on an interface in a regularly spaced array, and it seems improbable that they are nucleated by dislocations or other defects.

Regular cellular structures of the type shown in Fig. 14.18 are not related directly to the crystallographic orientation, but some crystallographic effects are observed. In particular, for h.c.p. structures with the growth direction almost in the basal plane, the pattern on the surface changes to parallel furrows and the three-dimensional substructure of the crystal is a set of laminar cells instead of a set of hexagonal rods. Rather similar structures have been observed in cubic crystals in the early stages of breakdown when elongated cells are frequently observed. Morris and Winegard found that nodes form the first segregated structures when the growth direction is near  $\langle 100 \rangle$  and  $\langle 111 \rangle$ , and elongated cells form for orientations near  $\langle 110 \rangle$ . The cell boundaries tend to lie on a plane in the  $\langle 100 \rangle$  zone and to be roughly symmetrical about  $\{111\}$  planes. It is suggested that this indicates that the most favoured perturbation is one in which the sides are bounded by the  $\{111\}$  planes which have lowest energy and are slowest growing. This produces long grooves on a  $\{110\}$  interface and depressions bounded on four sides on a  $\{100\}$  interface. Similar experiments have been made on aluminium alloys (Biloni, 1961; Morris and Winegard, 1969b).

Cellular structures also form in materials such as germanium which have relatively high entropies of fusion so that the solid-liquid interface develops facets. Bardsley *et al.* (1962) worked with gallium-doped germanium grown by the Czochralski technique and found that under conditions of constitutional supercooling, cells developed for any interface orientation except  $\{111\}$ . The non-planar interfaces are initially smooth but develop facets when parts of the interface reach  $\{111\}$  orientations. Elongated cells with a roof-top structure of two  $\{111\}$  facets are formed when growth is in a  $\langle 110 \rangle$  direction and regular cells shaped like pyramidal projections and depressions composed of four  $\{111\}$  facets result from growth in a  $\langle 100 \rangle$  direction. These observations are related to those of Morris and Winegard on metals, which do not develop facets but show a similar crystallographic dependence of morphology on orientation.

The experiments of Bardsley *et al.* also constitute the first experimental study of the breakdown of planar growth under conditions of convective mixing. Similar results were obtained by Hunt *et al.* (1968).

As already emphasized, the constitutional supercooling criterion is a thermodynamic condition and a true theory of interface stability must deal with the kinetics of the breakdown. The stability theory developed by Mullins and Sekerka (1963, 1964) and Sekerka (1965, 1967) is much more comprehensive and provides a description of the time evolution of the shape of a perturbed interface and of the accompanying temperature and composition fields. This theory was described on pp. 491–494, but was there applied only to the case of diffusion-controlled growth in the solid state. When growth from the liquid is considered, it is necessary to include the relevant equations for heat transport in both solid and liquid phases. In the simplest version of the theory, the assumptions are local equilibrium at the interface, isotropic surface free energy and absence of convection in the liquid. In the “time-independent” version of the theory, the steady-state solutions for the concentration and temperature, as given, for example, by eqns. (67.3)–(67.6a and b) for a planar interface, are tested for stability against an arbitrary small perturbation of the interface. This procedure is justified by the argument that perturbations with a

wavelength only slightly longer than the critical wavelength for instability will increase in amplitude very slowly and hence can be considered to be accompanied by steady-state diffusion fields.

Because the problem is linear for small perturbations, an arbitrary shape fluctuation may be Fourier-analysed, so that it is sufficient to treat a sinusoidal fluctuation in the planar interface, as postulated in eqn. (54.28). It was pointed out on p. 496 that the essential instability of the diffusion field (Fig. 11.3) is opposed by the stabilizing effect of the surface free energy. Mullins and Sekerka measured this tendency by a dimensionless parameter, here called  $A_{MS}$ , which is given by

$$A_{MS} = \left[ -i^2 v^s \sigma^{sl} \Upsilon T^{ls} \right] / [(1-i) \Delta h^{sl} D^l m c_0] \quad (67.17)$$

If  $A_{MS}$  is greater than unity, instability is impossible, but with typical experimental conditions  $A_{MS}$  may be smaller than  $10^{-3}$ . The condition for a planar interface to be unstable, i.e. for  $\dot{\delta}/\delta$  to be positive for some value of  $\omega$  in eqn. (54.28), is then found to be

$$(dT/dx)_e \leq m S_{MS} \quad (67.18)$$

where  $(dT/dx)_e$  is the effective temperature gradient and  $S_{MS}$  is the Mullins and Sekerka stability function which depends on  $A_{MS}$  and is given in graphical form by Sekerka (1965) as a function of  $\log(A_{MS})$  for  $i$  varying from 0.01 to 1.000. When the surface free energy is zero  $S_{MS} = 1$  and for other conditions  $0 < S_{MS} < 1$ , but it is unlikely that  $S_{MS}$  will be less than 0.8 if  $i$  is less than one.

The instability condition (67.18) is expressed in terms of the concentration gradient and both temperature gradients, and is thus difficult to compare directly with the condition for constitutional supercooling. However,  $(dT/dx)_l$  and  $(dc/dx)_l$  may be eliminated by the use of eqns. (67.7) and (67.18) to give an instability condition

$$(dT/dx)_l / \Upsilon + [\Delta h^{sl} / 2 v^s K_T^l] \leq - \left[ m c_0 (1-i) (K_T^s + K_T^l) / 2 D^l i K_T^l \right] S_{MS} \quad (67.19)$$

in terms of the temperature gradient in the liquid alone. The condition (67.19) differs from the constitutional supercooling condition, eqn. (67.16), by the inclusion of the extra (second) term on the left, by the factor on the right which depends on the relative thermal conductivities and finally by the function  $S_{MS}$ , also on the right. These modifications are not appreciable for most of the experiments which have been made on interface breakdown in metallic systems. For example, the latent heat correction varies from 30% to 3% of  $(dT/dx)_l$  in the experiments of Walton *et al.* on tin, described above, whilst the other two correction factors are about 1.5 and slightly less than one respectively. According to eqn. (67.19), a plot of  $(dT/dx)_l$  against  $c_0$  should have appreciable curvature only close to the origin, where it should give a finite intercept on the  $c_0$  axis.

The latent heat term which tends to stabilize the interface is appreciably larger for non-metallic materials, e.g. water, but so also is the ratio  $K_T^s / K_T^l$  (equal to four for water) which has the opposite effect. When this ratio is large, the condition for instability is much more dependent on the thermal gradient in the solid than on that in the liquid. In general,

the Mullins and Sekerka theory predicts that instability may exist without constitutional supercooling, or that constitutional supercooling may exist without instability, but in practice it is difficult to separate eqn. (67.19) from eqn. (67.16).

The time-dependent theory of interface stability was developed by Sekerka (1967). The steady-state solution is not taken as an unperturbed condition and the governing equations are the three-dimensional form of eqn. (67.8) in which the right-hand sides are set equal to  $(\partial T/\partial t)$ . The boundary conditions are linearized and the equations solved by a method which involves a Laplace transform with respect to the time and a Fourier transform with respect to the spatial coordinates. This theory enables the evolution of a particular fluctuation, or a Fourier component of a general fluctuation, to be followed as a function of time, at least until the fluctuation has become so large that the linearized equations may no longer be assumed to be valid.

The basic result of the time-dependent theory is an expression for the Laplace transform of a particular Fourier component of the spatial perturbation. Analytical inversion of this Laplace transform is not possible in general, but Sekerka treated some particular examples. For a system just on the verge of instability, the Mullins–Sekerka condition [eqn. (67.22)] is re-obtained, thus justifying the use of the steady-state approximation. In the limit of very long times, even if the complete steady state has not been attained, the use of a thermal steady-state approximation (almost always valid) leads to the same result. A modified stability criterion is predicted in the short time limit, so that perturbations which initially grow or shrink do not necessarily continue to do so. Sekerka showed, however, that this initial condition applies only to very short times.

Delves (1965) considered stability by perturbing the diffusion field rather than the interface shape, and he used different mathematical techniques, but his results are essentially equivalent to those of Mullins and Sekerka. In a second paper, Delves (1968) discussed stability when the melt is efficiently stirred, and concluded that a modified constitutional supercooling criterion proposed by Hulme is a good approximation.

The complete expression for the shape evolution of the component (54.28) of a perturbation in the nearly stable situation is an expression for  $\dot{\delta}/\delta$  which is considerably more complex than eqn. (54.32). Values of  $\dot{\delta}/\delta$  are negative at small wavelengths (large  $\omega$ ) and at large wavelengths (small  $\omega$ ) and are positive (if at all) at intermediate wavelengths, with a maximum at some value of  $\omega$ . As explained on p. 497, the negative values at small wavelengths are attributable to the stabilizing effects of surface free energy. At long wavelengths, the stability arises from the dominance of the effects of thermal gradients because solute redistribution becomes negligible. For typical growth conditions, the wavelength corresponding to the maximum value of  $\dot{\delta}/\delta$  is about 50  $\mu\text{m}$  and this is of the same order as the observed cell size.

When condition (67.19) is satisfied, the planar interface changes into a cellular surface which, over some range of growth rates and temperature gradients, seems experimentally to be stable against further fluctuations. The diffusion fields around non-planar interfaces are complex, and an exact solution for the shape presents formidable difficulties. Although the growth is quasi-steady-state, the departure of the interface from the planar shape means that the theory is less tractable than is the superficially similar model of eutectic

growth (see Section 55 and below). Moreover, in a full treatment, the effects of diffusion in the solid would have to be included. For the tips of the cells, the problem is similar to that of dendritic growth under an imposed temperature gradient (Burden and Hunt, 1974a and b) but diffusion in the grooves must also be considered. Tiller (1958) made calculations which showed that the grooves are very deep under typical growth conditions, and this has been amply confirmed by experiments in which longitudinal sections of a quenched, partly solidified specimen are examined (see, e.g., Sharp and Hellawell, 1971) and also by direct observation of the interface in solidifying transparent organic materials, used as analogues for metals (see, e.g., Jackson and Hunt, 1965). Figure 14.18 shows a typical cellular interface in a Pb–0.1% Sn alloy. For  $i < 1$ , there is experimental evidence to suggest that, even in dilute alloys, the solute concentration in the liquid at the root of a groove rises to its maximum possible value which might, for example, be the composition of a eutectic.

An obvious parameter to be related to the processing variables is the cell spacing and one possibility is that the observed spacing corresponds to the fastest growing perturbation. Early experimental measurements of spacing are, however, now considered to be mostly unreliable as Sharp and Hellawell (1971) showed that, especially at small velocities, appreciable coarsening of a cell structure takes place behind the solidification front and, in many cases, it is the final coarsened structure which has been examined. Sharp and Hellawell's own measurements indicated that the initial cell spacing does not vary appreciably with the freezing rate or the composition, but decreases with the imposed temperature gradient, possibly as  $(dT/dx_1)^{-1/2}$ . Also, the radius of curvature at the tip was reported not to be proportional to the spacing, but to vary as  $(dT/dx_1)^{-1}(\gamma c_0)i^{-1/2}$ . If growth conditions are changed, the simple cell structure remains stable until the tip radius reaches one-half of the spacing, i.e. the tip becomes approximately hemispherical. Further decrease in radius then leads to a fairly sharp morphological change in which simple cells change into branched forms.

There is considerable difficulty in defining the cellular to dendrite transition and its dependence on various growth parameters. The branched-cell structure reported by Sharp and Hellawell may represent an intermediate stage in this transition but, in other cases, there appears to be a more gradual change. Dendritic growth is not easily defined in unidirectional solidification but, as the solute content or the velocity of the interface is increased, the cellular growth front which is determined by the heat flow direction may change gradually to a dendritic front which has its primary and secondary, etc. arms along crystallographic directions. According to Morris and Winegard (1969), simple cells with smooth cross-sections first begin to deviate towards a preferred direction of crystal growth, and then (in cubic metals) the structure changes into a cruciform shape (i.e. a flanged or Maltese cross structure), from which side-arms later develop. Brody and Flemings (1966) and Bower *et al.* (1966) gave a more general discussion not only of axially symmetric dendrites but also of various dendritic plates which may form. Primary plates contain primary and secondary arms, and are parallel to the main growth direction, whereas secondary plates are transverse to the growth direction and roughly parallel to the growth front.

Many attempts have been made to relate the transition from cellular to dendritic growth to specified growth parameters. Plaskett and Winegard (1960) and many subsequent workers suggested that the transition depends upon  $(dT/dx_1)/\Upsilon^{1/2}$ , but others showed that the available data do not fit this relation.

An approximate theory of cellular growth may be based on the assumption that the cell spacing and solute distribution in the liquid so adjust themselves that constitutional supercooling is virtually eliminated over the whole of the intercellular region (Bower *et al.*, 1966). In the simplest treatment, transverse composition gradients are neglected and the liquid everywhere has a composition which is related to the local temperature by the liquid lines of the equilibrium diagram. This means that if isotherms are planes normal to the overall growth direction, the liquid is very nearly of uniform composition in any transverse section at fixed distance behind the cell tips. Moreover, the composition gradient parallel to the growth direction is given by

$$(\partial c/\partial x_1) = (\partial T/\partial x_1)/m \quad (67.20)$$

and conservation of solute at the tips requires that

$$\Upsilon(c_t - c_0) = -D^l(\partial c/\partial x_1)_t \quad (67.21)$$

where  $c_t$  is the liquid composition at the tips which are in a position  $x_1 = x_t$ . From eqns. (67.20) and (67.21)

$$c_t = \left[ 1 - \{D^l(dT/dx_1)/mc_0\Upsilon\} \right] c_0 \quad (67.22)$$

The solid composition at the tips is  $ic_0$ , so that an effective partition coefficient  $i'$  may be defined for the solid deposited at the tips where

$$i' = ic_t/c_0 = [1 - \{D^l(dT/dx_1)/mc_0\Upsilon\}]i \quad (67.23)$$

The cell tips will be at a temperature corresponding to the liquidus for composition  $c_t$ , and will be supercooled by an amount  $\Delta T^-$  from the liquidus temperature of the bulk composition,  $c_0$ , where

$$\Delta T^- = m(c_0 - c_t) = D^l(dT/dx_1)/\Upsilon \quad (67.24)$$

Equation (67.24) requires that the tip temperature for a particular temperature gradient increases with increasing growth rate, and early measurements of tip temperatures (Kramer *et al.*, 1963) supported this conclusion. Similarly, Sharp and Hellawell used microprobe analysis to measure the solute concentration deposited at the tips and found the predicted [eqn. (67.23)] variation of the effective partition coefficient with interface velocity  $Y$ . Later measurements of tip temperatures (Doherty *et al.*, 1973), though less readily interpretable, indicated an opposite effect. The experimental and theoretical work of Burden and Hunt (1974) resolved this difficulty; they showed that the supercooling decreased with increasing velocity at low velocities and high gradients, but increased with increasing velocity with high velocities and low gradients. The measured tip temperature thus shows a rather flat minimum when plotted against growth rate at constant gradient.

To develop a model of the growth, Burden and Hunt began with the usual assumption that the temperature varies only along the growth direction. The interface temperature at any point [cf. eqn. (66.5)] is then defined by

$$T_r^{\text{ls}}(c^{\text{l}}) = T_{\text{r}}^{\text{ls}}(c^{\text{l}}) - \Delta T_{\text{r}} - \Delta T^{\text{B}} \quad (67.25)$$

where  $T_{\text{r}}^{\text{ls}}(c^{\text{l}})$  is the liquidus temperature for composition  $c^{\text{l}}$  of the liquid at the interface,  $\Delta T_{\text{r}}$  is the deviation from this temperature to produce equilibrium at the tip of radius  $r$  and  $\Delta T^{\text{B}}$  is the usually negligible deviation from equilibrium needed to drive the interface forward at a constant speed. This equation may be expressed more succinctly by writing  $T_{\text{r}}^{\text{ls}}(c_0) - T_{\text{r}}^{\text{ls}}(c^{\text{l}}) = \Delta T^-$  and  $T_{\text{r}}^{\text{ls}}(c_0) - T_{\text{r}}^{\text{ls}}(c^{\text{l}}) = \Delta T_{\text{D}}$ , so that eqn. (67.25) becomes

$$\Delta T^- = \Delta T_{\text{D}} + \Delta T_{\text{r}} + \Delta T^{\text{B}} \quad (67.26)$$

This equation is just an expression of near equilibrium at the tip of a growing projection, and is similar to eqn. (66.5), but the detailed interpretation is different. Because the supercooling is constitutional, the temperature in the liquid rises continuously from the tips, and the lowest temperatures occur at the interface where the supercooling is  $\Delta T^-$ . In contrast, the temperature is lowest in the body of the liquid when a pure metal solidifies dendritically, and  $\Delta T^-$  in eqn. (66.5) gives the bulk supercooling. The specification of the interface temperature in terms of the composition  $c^{\text{l}}$  is completed by the equations

$$\Delta T_{\text{D}} = m(c_0 - c^{\text{l}}) \quad (67.27)$$

$$\Delta T_{\text{r}} = mc^{\text{l}}\Gamma^{\text{s}}(r_1^{-1} + r_2^{-1}) = \theta(r_1^{-1} + r_2^{-1}) \quad (67.28)$$

$$\Delta T^{\text{B}} \cong 0$$

where  $\Gamma$  is the Gibbs-Thomson coefficient introduced on p. 183. In dilute solutions,  $\Delta T_{\text{r}}$  is given to a good approximation by taking  $\theta = v\sigma^{\text{sl}}/\Delta s^{\text{sl}}$ , using parameters appropriate to the pure solvent as in eqn. (66.9). A more accurate expression is

$$\theta = v\sigma^{\text{sl}}/[ (1 - x^{\text{s}})(S_{\text{A}}^{\text{s}} - S_{\text{A}}^{\text{l}}) + x^{\text{B}}(S_{\text{B}}^{\text{s}} - S_{\text{B}}^{\text{l}}) ] \quad (67.29)$$

where  $x^{\text{s}}$  is the atomic fraction of component B in the solid phase and  $S_{\text{A}}^{\text{s}}$ ,  $S_{\text{A}}^{\text{l}}$ , etc. are the partial entropies of A and B [cf. eqn. (22.37)].

The liquid in front of the growing tips is assumed to have an average composition  $\bar{c}$  and an additional term  $\delta c$ . The average composition obeys eqn. (67.2) and thus decays exponentially away from the plane of the tips according to eqn. (67.3), giving for the gradient at a tip

$$(dc/dx_1)_t = \Upsilon(c_0 - \bar{c}_t)/D^{\text{l}} \quad (67.30)$$

and from eqn. (67.26)

$$\Delta T_{\text{r}}^- = (mD^{\text{l}}/\Upsilon)(dc/dx_1)_t - m\delta c_t + \Delta T_{\text{r}} \quad (67.31)$$

Burden and Hunt next made the same assumption as that of Bower *et al.* and Sharp and Hellawell for the average liquid composition behind the tips, which leads to

$$(dc/dx_1)_t = (dT/dx_1)/m \quad (67.32)$$

This is a reinterpretation of eqn. (67.20) which was formerly supposed to apply to the total concentration  $c$ .

The average composition gradient of eqn. (67.32) is just sufficient to take away solute from (or supply it to) a planar interface, as shown by eqn. (67.3); for the non-planar interface, the additional gradient of  $\delta c$  which is normal to the local surface element near the tip (and hence not parallel to  $x_1$ ) is needed to carry away solute rejected at the tip. As a first approximation, Burden and Hunt used the Zener assumption that the effective diffusion distance is of the order of the tip radius (see Section 54) and, after some manipulation, this gives

$$\Delta T^- = (D^l/\Upsilon)(dT/dx_1) - (mr\Upsilon/D^l) \cdot (1 - i)c_0 + 2\theta/r \quad (67.33)$$

The first term in this equation corresponds to the approximation previously used [eqn. (67.24)]; it is independent of tip radius and is completely dominant at high temperature gradients and low velocities. The second term, which is also positive, arises from the additional radial concentration gradient near the tips. The third term represents the effects of curvature. As in almost all other growth theories, there are still two undetermined quantities in the equation; with an imposed temperature gradient and growth velocity, these are the interface supercooling  $\Delta T^-$  and the tip radius  $r$ . The theory becomes determinate if it is assumed that  $r$  adjusts to give minimum undercooling, and this corresponds to the assumption of maximum velocity in free growth theories.

The theory was further elaborated to incorporate the Trivedi analysis of dendritic growth near the tip, and Hunt (1979) later showed how the theory could be used to predict the dependence of cellular or interdendritic spacing on growth parameters. The available experimental results at that time seemed to give reasonably good agreement with the predictions but, because of the uncertain effects of coarsening, comparison with experiment is difficult. However, over most of the growth range the first term is expected to dominate and under these conditions the spacing  $y_1$  is given by

$$y_1 \propto \Upsilon^{-1/4}(dT/dx)^{-1/2} \quad (67.34)$$

Hunt suggested that this relation should apply to both cell and dendrite spacings.

Despite its success in predicting some experimental results, however, the approximation and assumptions necessarily made in the Burden–Hunt and other models in order to obtain a solution are altogether too drastic, and once it was accepted that freely growing dendrites do not correspond to Zener's maximum growth rate hypothesis, the use of an equivalent optimizing condition for the growth of cellular or dendritic arrays no longer seemed reasonable. As already indicated for free dendrites, these considerations led Hunt and his students to develop numerical models for array growth in which all important interactions could be included.

The analytic theory for free growth into an alloy melt using the Ivantsov model with the tip selection criterion (66.14) is considerably more complex than that for pure metal dendrites (Trivedi and Kurz, 1994) and will not be described here in detail. The results show that with an assumed value of  $\sigma^* = 0.01$ , the dendrite tip velocity increases with composition to a maximum value and then declines again, and the tip radius passes

through a minimum as the composition is varied. This is in agreement with experimental results for the succinonitrile–acetone system (Chopra *et al.*, 1988); the maximum in velocity disappears at high undercooling.

For directional solidification of an alloy, the temperature gradient is externally superimposed, so that the temperature at the dendrite tip is determined mainly by solute diffusion and is related to the tip composition by the phase diagram modified for the capillarity effect. Again the full equation is complex but, if the temperature gradient effect is negligible, it simplifies to

$$\Upsilon r^2 = (\Gamma' D / \sigma^* k \Delta T_0) (c_0 / c^l) (1 / \zeta_0) \quad (67.35)$$

The last two brackets on the right of this equation are both approximately unity for dendritic growth at low velocities, so that the theory predicts under such conditions that  $\Upsilon r^2$  is constant, independent of velocity. This has been verified experimentally by directional solidification of succinonitrile–acetone (Esaka and Kurz, 1984) and the results also yield an experimental value of 0.02 for  $\sigma^*$ . The same value of  $\sigma^*$  was found in other systems (Trivedi and Mason, 1991), again casting doubt on eqn. (66.18).

Using the experimental  $\sigma^*$ , theoretical predictions of tip radius and temperature as a function of velocity can be compared with experiment. The comparison shows good agreement; nevertheless, it seems likely that the numerical methods of Hunt and his collaborators will ultimately prove superior to the very complex analytical models.

An outline description of Hunt's numerical approach was given in Section 66. Originally intended as a model for cellular growth, he found it could be applied also to dendritic growth. The model describes both steady-state and non-steady-state growth but neglects latent heat, and diffusion in the solid, and it assumes that the thermal conductivities of solid and liquid are equal. The solute transport problem in the liquid is solved by using a time-dependent finite difference method and the heat flow is included by assuming a moving linear temperature field. The liquid–solid interfacial free energy is included in anisotropic form and a steady state is obtained by allowing the time step to become infinite.

Hunt and Lu (1996) have fitted their numerical results to analytical expressions for the dendrite solute and dendrite curvature undercooling and for the dendrite spacing as functions of the reduced temperature gradient and the interface velocity. The equations are given in terms of the following dimensionless variables:

$$G' = G \Gamma i / (\Delta T_0)^2 \quad \Upsilon' = \Upsilon \Gamma i / D \Delta T_0 \quad (67.36)$$

where  $\Delta T_0$  is the supercooling for planar growth given by  $\Delta T_0 = m c_0 (i - 1) / i$ . In terms of these quantities, the solute supercooling at the interface is given by

$$\Delta T_s / \Delta T_0 = \{(G' / \Upsilon')^c + a (Y')^b + (1 - a) \Upsilon'\}^{1.2b} \quad (67.37)$$

where  $a = 8.734 + 5.931 \log_{10}(i) + 0.2578 [\log_{10}(i)]^2$  and  $b$  and  $c$  are given by similar complex expansions of  $i$ . The expression for the curvature supercooling is

$$\Delta T_r / \Delta T_0 = 0.41 (\Upsilon' - G')^{0.51} \quad (67.38)$$

and the half spacing,  $\lambda$ , of the dendrites is given by

$$\lambda \Delta T_0 / \Gamma i = 0.07798 (\gamma')^{d-0.75} (\gamma' - G')^{0.75} G'^{-0.6028} \quad (67.39)$$

where  $d$  is another function of  $i$ . The expressions are stated to be valid for  $G' > 10^{-10}$  and  $0.068 < i < 0.698$ ; this correspondingly limits the range of values of  $a$ ,  $b$ ,  $c$  and  $d$ ; for example,  $2 < a < 8$ . The equations illustrate the complexity of the numerical calculations and the virtual impossibility of replacing them with a satisfactory analytic theory.

The whole of this section has been concerned with the problem encountered in the solidification of single-phase alloys, and it is now appropriate to discuss the conditions under which two solid phases form simultaneously from the liquid. There have been very many experimental investigations of eutectic freezing, but comparatively little is known about peritectic or monotectic reactions.

Examination of typical microstructures of eutectic alloys shows that many morphological forms may be encountered, and classification of the structure is difficult, especially as the morphology of a particular alloy may vary considerably with the growth conditions. Early work on the classification of eutectic structures, notably by Scheil (1946), led to the concepts of normal and abnormal eutectics and also to the idea of the eutectic grain. Scheil regarded normal eutectic microstructures as the lamellar or fibrous forms in which simultaneous growth of the two solid phases occurs at a common interface to produce structures which in idealized form consist of alternate parallel lamellae of the two phases or parallel rods (fibres) of one phase regularly stacked in a matrix of the other. In controlled normal unidirectional freezing, a single set of lamellae or rods is produced, and may be regarded as the two-phase equivalent of a single crystal. When freezing begins at a number of different centres, as in a cast ingot, each nucleus gives rise to a differently orientated set of lamellae or rods, and the surfaces of mutual impingement of these sets will be clearly shown in the microstructure of the resulting solid. By analogy with single-phase solidification, each region of growth from a common centre is called a eutectic grain. It was established in early work (Straumanis and Brakss, 1935; Ellwood and Bagley, 1949) that all the lamellae or fibres of a given phase have essentially the same crystallographic orientation in a particular eutectic grain.

Scheil regarded as an abnormal eutectic microstructure any intimate mixture of the two phases which does not have the lamellar or fibrous form because the two phases for some reason are unable to grow cooperatively from the melt at the same velocity. A stable regular growth form is then impossible, but a wide variety of structures results from the free growth of the faster solidifying phase followed by crystallization of the second phase from the interdendritic liquid. The concept of a eutectic grain is then not so useful and different regions of primary growth cannot always be readily distinguished.

The morphological development of a particular eutectic may change with growth conditions, but following Hunt and Jackson (1966) it is now usual to divide eutectic systems into three categories by considering whether the solid-liquid interface developed by each phase of the eutectic mixture is faceted or non-faceted. As discussed on p. 650, an atomically rough, non-faceting interface is expected if the entropy of fusion is small, and

this includes most metallic phases with the exception of some intermetallic compounds. Interface facets develop, however, during the solidification of elements such as silicon and germanium, and of many molecular crystals. It follows that all metal-metal and many metal-intermetallic eutectics are in Hunt and Jackson's first group of non-faceting/non-faceting combinations of phases and there are a few organic and inorganic eutectics of this type. Systems in the second group of non-faceting/faceting phases are mainly alloys of metals with non-metals and this group includes, in fact, the two eutectic systems which have been, up to now, of greatest industrial importance, namely cast irons and aluminium-silicon alloys. The third category of faceting/faceting combinations leads to the most irregular microstructures but will not be discussed further here as there are very few metallic eutectics of this type.

In non-faceting/non-faceting eutectic systems, the two phases usually grow together from the liquid to produce lamellar or fibrous microstructures with the plane of the lamellae or the axis of the fibres aligned roughly parallel to the growth direction, i.e. the direction of heat extraction. There is usually a crystallographic orientation relation between the phases. Various "ribbon" morphologies intermediate between lamellae and fibres are also found in some alloys. Although the fibrous phase appears from a two-dimensional section to be dispersed in a matrix, examination has shown that the individual fibres are generally interconnected and have thus grown by branching from a single nucleus. The individual plates of a lamellar eutectic are similarly interconnected in three dimensions so that any eutectic structure forming by coupled growth from the liquid probably consists of two interwoven single crystals. Very regular structures are obtained by the unidirectional solidification of eutectic alloys, and important improvements in mechanical properties may be obtained by such procedures.

Experiments have shown that at temperatures below the eutectic temperature  $T^E$  it is possible to obtain coupled growth of the two phases from a wide range of liquid compositions. The resultant solid has a typical eutectic morphology and the compositions of the individual phases are close to the equilibrium composition at the temperature of formation. When solidification begins above  $T^E$  or below  $T^E$  at compositions sufficiently far removed from the eutectic composition, one of the phases forms first as large primary crystals which grow until the liquid is sufficiently enriched in the component being rejected by the primary crystals to allow coupled growth of the two phases finally to occur. It may happen in extreme cases that the second solid phase is difficult to nucleate so that coupled growth is not obtained even in the final steps. The second phase then solidifies as a thin film surrounding the crystals of the primary phase. This situation is usually described as a divorced eutectic. The reason why coupled growth is generally preferred even for liquid compositions well removed from the eutectic has been explained in Chapter 11; the cooperative nature of the diffusion process enables the two phases to grow together at a faster rate than either can achieve separately.

Unidirectional solidification of pure binary eutectic alloys involves coupled growth with a quasi-planar solid-liquid interface, as discussed in detail below.

In a supercooled liquid alloy, where growth is three-dimensional, each eutectic grain usually grows as a spherical nodule with a smooth well-defined solid-liquid interface.

Within such a region, the rods or lamellae of the greater eutectic microstructure are arranged in a general radial pattern and are again to be regarded as part of the highly branched single growth unit. Dendritic growth of the eutectic aggregates is sometimes observed at very high growth velocities or in the presence of impurities. Impurities can also lead to the formation of a kind of substructure within the eutectic grain which is analogous to the formation of the cellular substructure of a slightly impure metal. This substructure arises if the partition coefficients between the two solid phases and the liquid both differ from unity in the same sense, so that there is a long-range accumulation or depletion of the third element in front of both solid phases. The situation is then similar to that described previously as constitutional supercooling; the planar interface will no longer be stable, and a cellular structure will develop with impurity-enriched or -depleted cell boundaries. The scale of this structure is normally much coarser than that of the interlamellar spacing; Hunt and Jackson (1966) have illustrated directly the non-planar interface produced by unidirectional freezing of such an alloy, using an impure transparent organic eutectic to facilitate observation. The subgrains of this cellular-like structure are generally known as "eutectic colonies".

Non-faceting/faceting eutectic growth is sometimes coupled in the same way as non-faceting/non-faceting growth but the degree of cooperation between the phases is not so great, orientation relations are unusual and a wider variety of microstructures are obtained by varying growth conditions. In unidirectionally grown aluminium-silicon alloys, Day and Hellawell (1968) found that three types of structure develop in different ranges of  $(dT/dx_1)/\gamma$ . At high values of this ratio (exceeding  $10^5 \text{ }^\circ\text{C s mm}^{-2}$ ), the growth front is planar and the two phases grow virtually independently by long-range diffusion over distances of the order of 1 mm to give a structure in which massive silicon particles appear to be embedded in the aluminium matrix. Despite the microstructural appearance, however, dissolution of the matrix showed that the silicon particles are all interconnected. At lower ratios of temperature gradient to growth rate, the planar aluminium growth front breaks down, and the silicon phase forms as rods in an approximately close-packed array and with highly developed  $\langle 100 \rangle$  fibre texture. This represents growth by a steady-state diffusion process, but the silicon rods were found to be uneven in cross-section and growth instabilities in which silicon and aluminium spread alternatively across the interface were encountered. As  $(dT/dx_1)/\gamma$  is further reduced, the structures become more complex, with the rods developing side-arms in the form of  $\{100\}$  plates or  $\{111\}$  corrugated sheets. It is not obvious that such structures can be described even approximately by a steady-state growth theory. Finally, above a critical growth rate of about  $2 \text{ mm s}^{-1}$ , the silicon grows in an irregular flake form with a  $\{111\}$  habit plane and heavy internal twinning. There is no pronounced texture, and the twinning allows rapid growth on re-entrant surfaces, so that the silicon can keep up with the matrix phase. This type of structure is that most characteristic of normal non-directional solidification in ordinary laboratory or foundry practice.

Grey cast iron consists primarily of the austenite-graphite eutectic, and is another important example of a non-faceted/faceted system. The graphite forms as flakes which are aligned approximately in the heat-flow direction, but not in a regular pattern. The

basic growth mechanism is edgewise extension of the basal planes of the hexagonal graphite, and each flake contains numerous sub-boundaries and twins which assist the growth, as in the aluminium–silicon eutectic. When nucleated in a supercooled liquid, the three-dimensional growth form of the grey iron is roughly as spherical nodules, similar to that of most metal–metal eutectics. More regular lamellar structures, akin to those obtained from metal–metal eutectics, are produced by the unidirectional freezing of white cast iron, i.e. a stable austenite–cementite eutectic.

Small quantities of impurities produce pronounced changes in the microstructures of some faceted/non-faceted eutectics, and this effect is generally known as “modification”. The addition of sodium to aluminium–silicon alloys produces a modified structure in which the scale of the silicon dispersion is very much reduced, and the resultant structure has often been described as “globular”. Scanning electron microscope studies by Day and Hellawell (1968), however, have shown that the silicon is actually present in a fine fibrous form, and the structure is similar to that found in very rapidly cooled, unmodified alloys.

The many theories of modification which have been proposed are classified by Hellawell (1973) into descriptions based on the Al–Si–Na equilibrium diagram, nucleation theories and growth theories. The most probable explanations at present are in terms of limitations of the silicon growth rate caused by absorption of sodium on the silicon–liquid interface (Day and Hellawell, 1968; Davies and West, 1963–4). At low growth rates in pure materials, the silicon leads the aluminium phase, but at high growth rates, or in modified materials, the faceted silicon phase is unable to keep up with the aluminium phase, the solid–liquid profile is consequentially changed, and the finer structure is formed. This theory supersedes an earlier theory by Thall and Chalmers (1949) in which the sodium was assumed to be preferentially absorbed at the solid–solid interfaces, but it is consistent with the general observations of modifying effects in other systems. It is not clear, however, why the alkaline and alkaline earth elements should be so efficient in comparison with other impurities, and especially why sodium should be so much more effective than any other known addition.

An equally important modified structure is produced in grey cast iron by additions of cerium or magnesium. The graphite, which normally has the basal plane of the hexagonal structure parallel to the plane of the flake, adopts a nodular or spheroidal morphology, and the resultant structures are known as spheroidal graphite (SG) cast irons (Morrogh and Williams, 1968). According to Hunter and Chadwick (1972) each nodule of graphite is nucleated from a small flake. A similar change of morphology can be induced in the graphite phase of the nickel–carbon eutectic simply by increasing the growth rate and scanning electron micrographs of the structures formed at increasing cooling rates (Lux *et al.*, 1969) show that the surfaces of the nodule consist of small regions of graphitic basal planes oriented parallel to the surface. Double and Hellawell (1969) pointed out that the spherulitic morphology implies an appreciable component of growth normal to the basal plane, and they have suggested that it may represent an extreme form of the tendency to twisting and bending possessed by flake graphite. Graphite has a very imperfect structure; the flakes are normally not single crystals but consist of a number of composite sandwich

layers separated by low energy stacking faults involving rotations about the  $c$  axis by amounts of  $13^\circ$ ,  $22^\circ$  and  $28^\circ$ . In addition, bending of the flake crystals during growth is accomplished by repeated twinning (Double and Hellowell, 1969, 1971) and by the formation of non-crystallographic tilt boundaries.

Double and Hellowell envisage that multiple twinning events about all possible axes produce a highly imperfect spherical shell containing a multitude of tilt and twist boundaries between individual crystallites. Expansion of the sphere by growth normal to the basal planes is then possible because of the many imperfections. Presumably, the cerium or magnesium additions have an effect similar to the imposition of a high growth rate in forcing the production of an imperfect structure of this type.

The mathematical theory of the growth of a non-faceting/non-faceting eutectic may be developed from the theory of duplex growth given in Section 33 with only slight modifications. In view of experimental results, the liquid composition  $c_0$  may differ from the eutectic composition  $c^E$ , although the eutectic growth generally occurs very close to the eutectic temperature  $T^E$ , and the composition of the liquid phase in contact with  $a$  and  $b$  lamellae never deviates much from  $c^E$ . This means that although the composition difference in the liquid ahead of the central points of the lamellae must be qualitatively similar to that shown in Fig. 11.6, the total variation in composition both parallel and normal to the interface must be small.

## 68. THE STRUCTURE OF CAST INGOTS

The effects described in the last two sections may now be related to the more practical problem of the structure of a casting. There are many different procedures which are used to make ingots or semi-finished or finished products, and almost all of these procedures involve complex, non-steady-state heat flow from the liquid to the mould and thence to the surroundings. Analytical solutions to the heat flow equations can be obtained only for castings which have simple geometrical shapes, and even in these cases rather drastic assumptions have to be made. When such solutions are available, estimates of important kinetic variables such as the time of solidification, the extent of the "mushy" (solid-liquid) zone and the cooling rate of the solid may be obtained. The final metallurgical structures, and hence properties, are obviously very dependent on these kinetic parameters. This aspect of solidification is treated in some detail by Flemings (1974); here attention is confined to an outline discussion of the structure of a typical cast ingot in which the heat flow is radial.

The outside of the ingot (see Fig. 14.19) contains a layer of very fine "chill" crystals. Succeeding this layer are long columnar crystals, growing perpendicular to the mould walls into the interior of the casting. In the centre of the ingot, there are often roughly equi-axed grains and if macroscopic shrinkage cavities or gas blow holes are formed, they are located in the centre of the upper part of the ingot, which solidifies last. The above description is in general terms only, and the details naturally differ with the individual metal or alloy, the pouring and mould temperatures, the nature of the mould, the size of

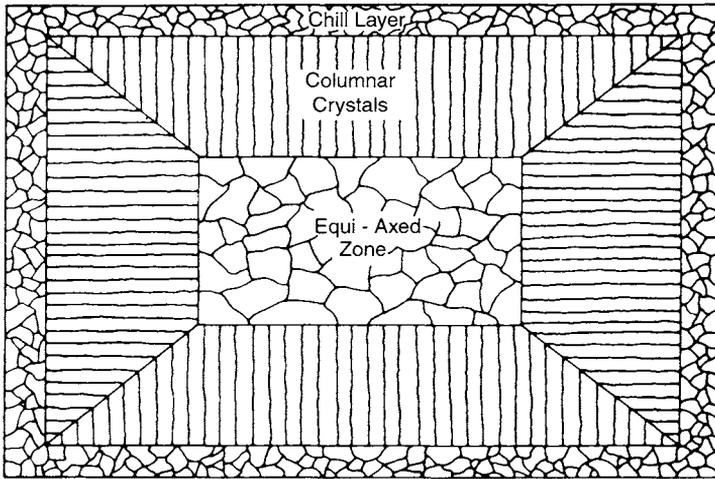


FIG. 14.19. The structure of a cast ingot (schematic) (after Hensel, 1937).

the ingot, etc. Three distinct zones are not always present; for example, the central region of equi-axed crystals is not found in ingots of very pure metals, and the outer layer of chill crystals can be eliminated by casting into a hot mould.

Some of the most complete experimental investigations into the structures of cast ingots appear to be those of Northcott (1938, 1939, 1946). He showed that the columnar crystals are often dendritic and that they have a preferred orientation such that (for cubic materials) a  $\langle 100 \rangle$  direction is parallel to the growth direction. The equi-axed crystals at the centre of alloy ingots are always dendritic and have random orientations. The extent of the columnar zone in alloy ingots varies with the separation of the liquidus and solidus curves.

Immediately after pouring, the liquid in contact with the walls of the mould will be cooled to below its freezing point and solid will be nucleated homogeneously or heterogeneously. In normal conditions, the supercooling will be appreciable and nucleation will be rapid. The large number of nuclei formed grow into the layer of chill crystals and will have random orientations; there may be some preferred orientations if the walls of the mould act as sites for heterogeneous nuclei. The number and size of the chill crystals will vary with the initial cooling rate, and metal moulds will produce smaller and more numerous crystals than will those of lower thermal conductivity. If the mould is preheated to a temperature approaching the freezing point, the solidification will be slower, and the few nuclei which are formed will grow into larger crystals. There is then no visible chill layer. Similar effects may also occur in ingots, where the chill crystals which form first are melted again by the hotter liquid so that solidification eventually resumes again with much smaller supercooling as the mould walls have then been heated.

The formation of nuclei will cease when there is a more or less continuous solid layer around the outside of the ingot. Heat withdrawal from the solidifying liquid must then

take place through this solid layer and the whole of the liquid must be at a temperature above the temperature of the solid-liquid interface. In a pure metal ingot, this means that the liquid cannot be supercooled and no further nuclei will be formed. Solidification proceeds by the growth of existing crystals inwards from the walls to produce the columnar zone, and continues until crystals from opposite surfaces meet in the centre. The columnar crystals will consist of those chill crystals which are most favourably orientated for rapid growth, the remaining chill crystals being crowded out. The columnar zone thus shows preferred orientation so that the most rapid growth direction coincides with the direction of the maximum temperature gradient normal to the mould walls. Note that, because the growth is non-dendritic, the liquid-solid interface must be nearly planar.

Northcott's results, and those of other workers, indicate that the columnar crystals in cubic metals always tend to have a  $\langle 100 \rangle$  direction parallel to the direction of rapid growth. This raises difficulties in connection with the above description as a preferred  $\langle 100 \rangle$  direction is associated with dendritic growth and there is some evidence that the tip of the dendrite is formed from four sets of  $\{111\}$  planes. Whatever the detailed mechanism of non-dendritic growth, the preferred orientation in the columnar zone might have been expected to be related to the primary growth form. In fact, Rosenberg and Tiller (1957) have shown that when zone-refined lead is solidified in a mould, the preferred axis along the growth direction is  $\langle 111 \rangle$ . They used a water-cooled mould, arranged to give unidirectional solidification, and they verified that the growth front was planar. All crystals observed were nucleated on the bottom face of the mould, but 80 of them were crowded out in the subsequent growth.

Rosenberg and Tiller found that the addition of  $10^{-4}$  wt.% of silver to the zone-refined lead gave a grain structure which was almost random in orientation, although there was a slight preference for a  $\langle 100 \rangle$  direction. This concentration of solute was insufficient to produce a stable cellular interface, but the planar interface was beginning to be unstable. Addition of  $5 \times 10^{-4}$  wt.% silver gave a stable cellular interface and the orientations were found to have a pronounced  $\langle 100 \rangle$  texture. The characteristic  $\langle 100 \rangle$  texture in the columnar zone is thus produced only when there is sufficient impurity present to give a stable cellular interface, i.e. a Smialowski substructure. However, later results do not support this conclusion.

Somewhat different results were obtained by Walton and Chalmers (1959) who cast lead and aluminium under conditions of cellular or dendritic growth and observed a texture only with a dendritic interface. They concluded from this that competition between grains of different orientations is effective only when the growth is dendritic. Later work by Hellawell and Herbert (1962) confirmed that a cellular interface does not generally give any pronounced texture, but clearly established for a range of materials and crystal structures that textures are produced by both planar and dendritic structures. Their correlation between the interface morphology and the texture is summarized in Table XII, and shows that in lead, zinc and bismuth the preferred orientation produced by planar growth differs from that produced by dendritic growth. The texture of tin appears to be independent of the interface structure, and the magnesium could not be refined sufficiently

TABLE XII. RELATION BETWEEN INTERFACE MORPHOLOGY AND TEXTURE  
(HELLAWELL AND HERBERT, 1962)

Metal	Structure	Planes of close packing	Interface Morphology and Orientation Planes normal to growth axis or Directions parallel to growth axis		
			Planar	Cellular	Dendritic
Lead	f.c.c.	{111}	{111}	random	$\langle 100 \rangle$
Zinc	h.c.p. ( $c/a = 1.856$ )	{0001}	{0001}	random	$\langle \bar{1}210 \rangle$
Magnesium	h.c.p. ( $c/a = 1.62$ )	{0001}	—	random	$\langle \bar{1}100 \rangle$ or $\langle \bar{1}210 \rangle$
Tin	b.c. tetr.	{110}	{110}	{110}	{110}
Bismuth	Orthorhombic	{110}	{110}	—	$\langle 111 \rangle$
$\beta$ -brass	b.c.c.	{110}	—	—	$\langle 100 \rangle$

to give a planar interface. The  $\beta$ -brass alloy was chosen to be representative of a b.c.c. structure but only dendritic interfaces were obtained.

In all cases, the results in Table XII are consistent with the rule that planes of closest packing are normal to the growth direction for planar interfaces and enclose the growth direction for dendritic interfaces. Hellawell and Herbert also found that when metals of different purities were cast under identical conditions, the growth rate was much smaller for planar interfaces than for dendritic interfaces.

More recent investigations of dendritic columnar growth (see, e.g., Kattamis and Flemings, 1966; Bower and Flemings, 1967) have shown the importance of dendritic plates which often extend over many primary dendrite arms. In cubic crystals, the primary dendrite plates form on the {100} planes which are most nearly parallel to the direction of heat extraction (Bower *et al.*, 1966). Dendritic plates are also observed in non-cubic systems, for example in tin-rich tin-bismuth alloys with a body-centred tetragonal structure (Ahearn and Flemings, 1967). Here the primary dendrite arms grow along [110] (see Table XII) and the secondary dendrite arms grow along  $[1\bar{1}\bar{1}][\bar{1}11]$  and [112], giving dendrite plates parallel to  $(1\bar{1}2)(\bar{1}12)$  and  $(1\bar{1}0)$  respectively. Single-phase alloys, or alloys in which there is only a minor amount of eutectic solidification, usually have well-developed plates, in contrast to alloys in which an appreciable fraction of the solid forms as a eutectic mixture.

An important result of dendritic columnar growth is the microsegregation of solute in directions transverse to the dendrite arms or plates. An approximate treatment of this microsegregation was given by Bower *et al.* (1966) who considered a volume element which has a small thickness in the growth direction ( $Ox_1$ ) and which extends from the centre of a cell or dendrite to a point midway to a neighbouring cell or dendrite. They solved the diffusion equation for this element assuming steady-state solidification and diffusion only in the liquid and obtained an expression for the composition of the liquid immediately adjacent to the solid. The variation of solute composition along the volume element is given by a slightly modified Scheil equation, but it should be noted that the segregation here is on a much finer scale within distances equal to the interdendritic

spacing rather than the length of the entire crystal as in normal growth. An important prediction is that the last interdendritic drops to solidify are very considerably enriched (or depleted) in solute. Thus if  $i$  remains constant until terminated by (say) a eutectic reaction, some eutectic will form in the interdendritic regions, even for very low initial concentrations of solute in the alloy. This prediction has been verified in many experiments. In aluminium-copper alloys, for example, the limit of solid solubility of copper in aluminium is  $\sim 5\%$ , and the eutectic composition is 35 wt.% copper, but some interdendritic eutectic is found for alloy compositions of  $< 0.5$  wt.% copper.

Measurements of microsegregation by electron microprobe analysis and other techniques have been made for various alloys (see, e.g., Kattamis and Flemings, 1966; Bower *et al.*, 1966). The results are qualitatively in agreement with the above theory and, for dendritic growth in practice, the ratio  $(dT/dx_1)/\Upsilon$  is so small that the effective distribution coefficient is effectively zero. Quantitative agreement is less good, the minimum composition observed being greater than that predicted and the volume fraction of austenite being less than that predicted. Flemings (1974) summarized the experimental evidence which indicated that the main reason for this discrepancy is the neglect of solid-state diffusion during and after solidification. In an approximate treatment of the problem for a plate-like dendritic morphology (Brody and Flemings, 1966; Bower *et al.*, 1966)

$$c^s = ic_0[1 - \zeta\{1 + (4D^s t/\gamma^2)\}^{-1}]^{i-1} \quad (68.1)$$

where  $\gamma$  is the arm spacing and where diffusion in the solid is also considered. This theory can be fitted to experimental results for the solidification of an aluminium-4.5 wt.% copper alloy.

Experimental results on the spacing of primary and secondary dendrite arms in castings are usually plotted as a function of the average cooling rate during solidification, which is given by the product of the temperature gradient,  $(dT/dx)$ , and the growth velocity,  $\Upsilon$ . According to Flemings (1974), most such measurements lead to a power law in which the spacing  $\gamma$  varies as  $\Upsilon(dT/dx_1)^{-n}$  with  $n \simeq 1/2$  for primary arms and  $n \simeq 1/2-1/3$  for secondary arms. However, these results are undoubtedly influenced by convection and by coarsening subsequent to the initial growth; this latter process is particularly marked for secondary spacings. The results for primary spacings do not seem to be inconsistent with Hunt's theory because, as noted above, it is difficult to distinguish between  $n = 1/2$  and  $n = 1/3$  in the growth rate dependence.

During growth, the primary dendrite spacing may be locally increased where necessary by elimination of one or more of the dendrites which lag behind the others if the spacing is too small, and similarly the primary spacing may be decreased by the growing forwards of a tertiary arm to fill a too large gap between two primary arms, and so in effect to become a new primary arm. Similar processes occur among secondary arms, and of particular importance is the instability which develops among some secondary dendrites far removed from the primary tips. The effect of constitutional supercooling is much reduced away from the primary tips, and secondary arms of small radii of curvature may become unstable and begin to melt again. The driving force for this remelting is excess surface energy and it leads to the coarsening mentioned above. However, it appears also to have

another important consequence in connection with liquid convection, namely that of grain multiplication. A secondary (or tertiary) arm which is slightly thinner near its root may be melted off there to leave a small free crystal immersed in the liquid. If this is then carried away into a supercooled region, it may grow into a large new crystal without the need for a separate nucleation event. A number of experiments have shown that this is often the source of the grains in the central equi-axed portion of an ingot. Convection not only serves to carry away the melted-off fragments to other portions of the liquid but also helps to dissipate the heat released; moreover, if the convection is turbulent, heat pulses are transmitted to the solid interface and may accelerate the melting-off process.

As might be expected from this description, convection has a large effect on the grain size of a casting and on the relative extent of the columnar and equi-axed regions. The grain size and the extent of the columnar zone both increase with increasing convection. It has also been suggested that the effects of heat flow are enhanced by mechanical breaking of dendrite arms as a result of the stresses caused by fluid flow. Whilst this may seem improbable, there is current interest in a deliberate mechanical process of grain refinement by agitating a solid-liquid slurry. Spencer *et al.* (1972) have shown for several alloys that vigorous agitation can prevent significant strength developing for slurries or mushy zones containing up to 40% solid, and it is interesting to note that this procedure rather resembles the age-old method of making high quality ice cream.

Commercial exploitation of the formation of a slurry of fine approximately spheroidal solid particles by vigorous agitation is now taking place, and two types of slurry casting procedures have been developed. In the first of these, known as "rheocasting", the liquid is vigorously agitated until it is 50% solid and is then fed from a continuous casting furnace into the shot chamber of a die-casting apparatus. The liquid-solid mixture is thus injected into the die to produce a high-quality casting. In the second process, "thixocasting", a rheocast ingot is first produced; this ingot is then sectioned into charges and each charge is reheated into the liquid-solid region to attain the required softness before being fed into the die-casting machine. The process, first described by Flemings *et al.* (1976), has been found to give adequate die life in steel die-casting experiments, and encouraging results on the structure and properties of thixocast steels have been obtained.

An earlier explanation of the equi-axed region of an ingot of an impure metal or an alloy (Winegard and Chalmers, 1951) was the progressive development of a constitutionally supercooled mass of liquid in the centre of the ingot as the solid columnar crystals grow inward and gradually change the composition of the remaining liquid. The constitutional supercooling accounts for the dendritic form of the columnar growth, but was eventually supposed to become so large that new crystals of random orientation are nucleated ahead of the columnar crystals. It is possible that such independent nucleation does occur in certain circumstances, but it is now clear that the origin of the equi-axed zone is usually the grain multiplication process described above.

Winegard and Chalmers' description clearly implies that in an alloy ingot there will be a macroscopic variation of composition, in which the outer layer will have a composition corresponding to the material of higher freezing point. This is known as normal segregation. Not infrequently, however, the overall composition changes in the opposite

sense to produce so-called inverse segregation. Inverse segregation is encountered most frequently in alloys which undergo a rather large volume change on freezing and it is ascribed to the solute-rich regions (solute-poor for  $i > 1$ ) of liquid which solidify last being sucked into interdendritic shrinkage cavities in the columnar zone.

In most ingots or castings, macrosegregation is much more complex and the composition does not vary monotonically from the outside to the centre of the ingot. Sectioned steel ingots, for example, often show lines or channels of segregation which are V-shaped along the centre line of the ingot and nearly vertical in the outer and upper parts of the ingot; these are known as A segregates. These segregates are solute-rich, and are associated with porosity, whilst a conically shaped region in the centre of the lower part of the ingot is solute-poor. Other alloys show different patterns of segregation, and the effects observed are now generally recognized as the results of various forms of mass flow of liquid and solid phases. In the early stages of solidification, for example, the free crystals which are formed by melting off of side-arms will then either sink downwards or float upwards in the liquid and this is one important source of macrosegregation. Movement of liquid in those parts of the ingot which have already partly solidified may also occur because of solidification shrinkage, thermal convection and differences of density or temperature induced by the release of latent heat. In some cases, instabilities can develop in this forced convection, where rapid flow of liquid induces local remelting of solid, which in turn increases the flow rate. Such instabilities are probably responsible for the concentration of segregates into localized channels.

The large number of variables and the need to consider both heat and mass flow make it difficult to develop a quantitative theory of macrosegregation and even a qualitative description of the segregation to be expected in given circumstances is difficult. A fuller description of this important field is given by Flemings (1974).

## 69. MELTING

Quantitative observations on the melting of bulk solids similar to those on the freezing of bulk liquid have never been made. The reason is that there is an essential asymmetry between the kinetics of the two processes; whereas all liquids may be made to supercool under suitable circumstances, the superheating of a solid is not observed under normal conditions. The transition solid-liquid is thus unique among phenomena which are classed as nucleation and growth reactions, as the existence of a metastable solid phase is not normally possible if the stable state is liquid. The theory of homogeneous nucleation cannot explain this; it predicts that the rate of formation of liquid nuclei should not become appreciable until the solid is superheated by an amount similar to the supercooling needed to freeze the liquid.

There are two possible explanations of the inability to superheat a bulk solid. One, based on the various lattice theories of liquids, assumes a near continuity of state between the solid and liquid. At the melting point, the long-range order of the lattice is assumed to break down spontaneously at all parts of the solid, so that melting is effectively a homogeneous process

without the need for thermally activated nucleation. Such theories are difficult to reconcile with the sharpness of the transformation, which is undoubtedly a first-order change.

The alternative explanation utilizes the conventional theory of heterogeneous nucleation. It is assumed that, under normal conditions, melting always begins at the free surface of the solid which is able to act as its own heterogeneous nucleating agent. A liquid droplet formed on the surface will have free energy terms corresponding to the liquid–solid and free liquid (or liquid–vapour) interfaces. If the sum of these is less than the free energy of the original solid interface which is eliminated, the curve of  $\Delta G$  against  $n$  will have negative slope for all  $n$  and there is no activation energy for nucleation. The condition for this was seen on p. 453 to be  $\theta < 0$ , i.e. the liquid wets the solid in the presence of the vapour. It is known that this condition is satisfied for most liquids in contact with solids of the same composition.

The nucleation theory implies that if liquid can be prevented from forming at the free surface, it should be possible to superheat the solid. In practice, grain boundaries or dislocations might be almost equally effective heterogeneous nucleation sites, so that nearly perfect single crystals are also required. Nucleation at the surface could be prevented either by maintaining it at a lower temperature than the interior of the solid, or by coating it with a coherent layer of some solid S of higher melting point. The coating should be some substance which catalyses the liquid–solid change so that the surface energy of liquid–S interfaces is greater than that of solid–S interfaces.

The surface coating technique has been applied to studies of embedded small particles produced by rapid solidification of monotectic systems, as described on p. 634, but to the author's knowledge there has been only one attempt to use the surface cooling method. Khaikin and Benet (1939) worked with tin specimens heated by a large electric current and cooled externally so that the centre of the specimen was 1.2 °C hotter than the surface. Polycrystalline specimens and defective single crystals were observed to begin to melt internally, but good single crystals began melting from the surface, thus indicating that the interior had superheated by up to 2 °C. This degree of superheating is very much smaller than the theoretically possible amount of 100 °C for homogeneous nucleation in tin, and some more convincing demonstration of solid superheating would be welcome.

Some workers have described experiments which seem to indicate that in polycrystalline materials the boundaries begin to melt at a lower temperature than the mass of the material. Chalmers (1940) applied a tensile stress across the grain boundary of a tin bicrystal and found that it separated along the boundary. In a number of experiments, he found that the temperature of separation was consistently 0.14 °C below the ordinary melting point. The depression was independent of the relative orientations of the lattices, except that no effect was observed with twin boundaries. Further observations were made by Chaudron *et al.* (1948) on high purity aluminium. They placed a plate of the material in a temperature gradient and observed the grain boundaries to melt ahead of the general solid–liquid interface, the lowering of melting point being estimated at 0.25 °C. Pumphrey and Lyons (1951) also used high purity aluminium with a tensile stress method; they reported grain boundary melting at a temperature as much as 4 °C below the normal melting point of the pure metal.

The interpretation of these experiments was rather confused for some time and there was much discussion as to whether the lower melting point is attributable to the segregation of impurities to the boundaries or else is an intrinsic property of the boundaries themselves. Chalmers believed his results to be due to genuine grain boundary melting, and he assumed that in any boundary some atoms possess a maximum strain energy which changes the effective melting point. It now seems certain that theories of this kind are incorrect. In principle, the presence of a grain boundary may lower the melting point of the whole specimen, but any localized change of melting point can only be due to non-equilibrium segregation of impurities. Arguments for this conclusion have been given in a very general form by Shewmon (1957), who pointed out that the condition for equilibrium of the solid is the equality of the chemical potential of the  $i$ th component in all parts of the specimen. The chemical potential has to be rather strictly defined, the number of lattice sites and the stress being held constant as well as the usual variables. When equilibrium is attained,  $g_i$  for each component is independent of position in the solid; this equilibrium may or may not involve differences in composition between grain boundaries and grain interiors. The melting point is defined by  $g_i^l = g_i^s$  and hence must also be independent of position in the solid. Thus, so long as equilibrium has been attained, solute segregation cannot lower the melting point of the boundaries, and any measured difference between the boundaries and the grain interiors must be due to non-equilibrium segregation.

Shewmon suggested that observed early melting of the boundaries might be due either to non-equilibrium segregation or more simply to rapid nucleation and growth in the boundary regions producing a false impression that the melting there begins at a lower temperature. Careful experiments by Weinberg and Teghtsoonian (1957) on the melting of tin and aluminium support this latter conclusion. They found that the boundaries parted at the melting point for misorientations greater than  $12^\circ$ ; for small angle boundaries, there was no tendency to part at the boundaries. Melting always took place at the melting point of the bulk material to within the accuracy of the measurements (0.02 °C). There was a time lag after reaching the melting point before a boundary parted. This time decreased with increase in the heating rate (i.e. in the rate of supply of latent heat), with increase in stress, with increased amount of impurities and with increasing orientation difference in the range  $11$ – $15^\circ$ . Assuming that nucleation of liquid occurs in the outer (free) surface of the bicrystal specimens, the delay time may represent the time taken for the liquid to penetrate right across the boundary. Parting at the boundary will be observed when the liquid grows along it more rapidly than into the crystals, and although a detailed theory has not been developed, the variation of delay time with the above parameters is of the type expected.

As pointed out above, any grain boundaries in a specimen will, in principle, lower the melting point of the whole specimen as will any departure of the external shape of a finite specimen from the equilibrium shape given by the Wulff construction. For macroscopic specimens, the extra free energy of internal or external surfaces is negligible, and the lowering is quite insignificant (of the order of  $10^{-4}$  °C). With very small specimens, however, an appreciable effect is possible and this may complicate the interpretation of

experiments on the melting of very small crystals. The effect was first noticed by Takagi (1954), whose experiments on thin films of lead, tin and bismuth were described in connection with solidification on p. 628. Using thin films, he found depressions of the melting point of 7–41 K for lead, 30 K for tin and 23 K for bismuth.

Subsequent experiments by other workers on the melting behaviour of small droplets confirmed that the temperature of melting was below the equilibrium melting temperature of bulk samples, and in most cases the reduction in temperature was proportional to the reciprocal of the particle size, as predicted by thermodynamic theory. Equating the free energy of a liquid droplet of radius  $r$  to that of a solid droplet and including the surface energy terms gives the equilibrium melting temperature of the droplet as  $T^{\text{sl}} - \Delta T$  where

$$\Delta T/T^{\text{sl}} = 3(\sigma^{\text{s}} - \sigma^{\text{l}})v'/r\Delta h^{\text{ls}} \quad (69.1)$$

Here  $T^{\text{sl}}$  is the bulk melting temperature,  $\sigma^{\text{s}}$  and  $\sigma^{\text{l}}$  are the free energies of the solid and liquid-free surfaces respectively,  $v'$  is the mean atomic volume and  $\Delta h^{\text{ls}}$  is the enthalpy of melting per atom. The droplet has a lower equilibrium melting temperature than the bulk metal provided that  $\sigma^{\text{s}}$  is greater than  $\sigma^{\text{l}}$ . The experimental results are consistent with reasonable values for the difference in surface energies. However, a different dependence of melting on particle size has been found for small particles embedded in a surrounding metal matrix.

Saka *et al.* (1987) confirmed by in-situ observations in a high voltage electron microscope that melting of indium particles embedded in splat-cooled thin foils of aluminium began from the indium aluminium interface and progressed towards the centre. Measurements of the temperature showed that melting took place at temperatures as high as 197 °C, although the normal bulk melting temperature is only 145 °C. The increased melting point was assumed by Saka *et al.* to be due to a displacement of the equilibrium melting temperature which is given by eqn. (69.1) with  $\sigma^{\text{s}}$  and  $\sigma^{\text{l}}$  being replaced by the free energies of the solid indium–aluminium interface and the liquid indium–solid aluminium interface respectively. Thus, the melting point could be increased if the energy of the liquid indium–solid aluminium interface is greater than that of the solid indium–solid aluminium interface.

Very similar results were obtained by Zhang and Cantor (1990), who studied dispersions of indium and lead in aluminium produced by melt spinning. The enclosed particles were 5–100 nm in diameter and had a cube–cube (or near cube–cube) orientation relation with the aluminium matrix; the particles had a truncated octahedral shape and were bounded by facets of  $\{111\}$  and  $\{100\}$  planes. Melting was recorded by a differential scanning calorimeter as well as by observation in a transmission electron microscope and temperatures up to 40 K above the bulk temperature were recorded. However, particles in contact with grain boundaries in the aluminium were found to melt at 1–7 K below the bulk melting temperature. Zhang and Cantor rejected the explanation given by Saka *et al.* on the grounds that it was improbable that the solid–solid surface free energy is less than the liquid–aluminium surface free energy and also because analysis of their results gave no justification for any substantial variation of melting temperature with particle size. Instead Zhang and Cantor attributed the high melting temperatures to a kinetic barrier to

nucleation; as suggested above, the catalytic effect of the free surface is prevented by the coherent coating.

In later experiments with cadmium enclosed in aluminium, Zhang *et al.* (1994) found only a depression of the melting temperature by about 7 K for 20 nm diameter particles and 9 K for 14 nm particles. It may be significant that in this case the particles did not have all interfaces as planar facets; there were two planar interfaces in which  $\{0001\}_{\text{Cd}}$  were parallel to  $\{111\}_{\text{Al}}$  but the other interfaces were curved. These results were interpreted as a genuine change in equilibrium melting temperature with particle size, and the authors also concluded that in this case there was no evidence of any kinetic barrier to nucleation of the liquid.

The evidence above and in other experiments that the melting is influenced by the nature of the interface between the particle and a constraining matrix has recently been directly tested by Sheng *et al.* (1996). They prepared two kinds of aluminium-embedded lead particles, one by melt spinning and the other by mechanical alloying (ball-milling fine powder mixtures). The melt-spun particles had the epitaxial relationships described above; the ball-milled specimens had random orientations and shapes, but a similar size range (5–50 nm, compared with 5–30 nm). In agreement with Zhang and Cantor, the melt-spun samples were found to melt at temperatures of 11–40 K above the bulk melting point whilst the other samples melted about 13 K below the bulk melting temperatures. The authors suggest that the free energy of the solid lead–solid aluminium interface may be greater than that of the liquid lead solid aluminium interface for the disordered interfaces of the randomly orientated particles but may be smaller for the coherent interfaces of the epitaxial particles. Alternatively, nucleation may be difficult on the epitaxial interfaces, but easy on the disordered interfaces, so that the apparent difference is a wholly kinetic effect.

## REFERENCES

- AHEARN, P. J. and FLEMINGS, M. C. (1967) *Trans. Metall. Soc. AIME* **239**, 1590.  
ATWATER, H. and CHALMERS, B. (1957) *Can. J. Phys.* **35**, 908.  
BARBIERI, A., HONG, B. C., and LANGER, J. S. (1987) *Phys. Rev. A* **35**, 1802.  
BARDSLEY, W., BOULTON, J. S., and HURLE, D. T. J. (1962) *Solid-State Electron.* **5**, 395.  
BECKER, R. (1937) *Z. Metallkd.* **29**, 245; (1938) *Ibid.* **32**, 128.  
BENSIMON, D., PELCE, P., and SHRAIMAN, B. I. (1987) *J. Phys. (Paris)* **48**, 2081.  
BILONI, H. (1961) *Can. J. Phys.* **39**, 1541.  
BILONI, H., BOLLING, G. F., and COLE, G. S. (1966) *Trans. Metall. Soc. AIME* **236**, 930.  
BOLLING, G. F. and TILLER, W. A. (1961) *J. Appl. Phys.* **32**, 2587.  
BOSWELL, P. G. and CHADWICK, G. A. (1980) *Acta Metall.* **28**, 209.  
BOWER, T. P. and FLEMINGS, M. C. (1967) *Trans. Metall. Soc. AIME* **239**, 216.  
BOWER, T. F., BRODY, H. D., and FLEMINGS, M. C. (1966) *Trans. Metall. Soc. AIME* **236**, 624.  
BRIDGEMAN, P. W. (1924) *Proc. Am. Acad. Arts Sci.* **60**, 305.  
BRODY, H. D. and FLEMINGS, M. C. (1966) *Trans. Metall. Soc. AIME* **236**, 615.  
BUELLER, M. J. (1934) *Z. Kristallogr., Kristallgeom., Kristallphys., Kristallchem.* **89**, 195.  
BURDEN, M. H. and HUNT, J. D. (1974a) *J. Cryst. Growth* **22**, 99; (1974b) *Ibid.* **22**, 109; (1974c) *Ibid.* **22**, 328.  
BURTON, J. A., PRIM, R. C., and SLICHTER, W. P. (1953) *J. Chem. Phys.* **21**, 1987.  
CECH, R. E. and TURNBULL, D. (1951) *Trans. Metall. Soc. AIME* **191**, 242.

- CHADWICK, G. A. (1962) *Acta Metall.* **10**, 1.
- CHALMERS, B. (1940) *Proc. R. Soc. London, Ser. A* **175**, 100; (1949) *Ibid.* **196**, 64; (1954) *Trans. Metall. Soc. AIME* **200**, 519.
- CHALMERS, B. and MARTIUS, H. R. (1952) *Philos. Mag.* **43**, 680.
- CHAUDRON, G., LACOMBE, P., and YANNAQUIS, N. (1948) *C. R. Acad. Sci.* **26**, 1372.
- CHOPRA, M., GLICKSMAN, M. E., and SINGH, N. B. (1988) *Metall. Trans.* **19A**, 3087.
- COLE, G. S. and WINEGARD, W. C. (1969) *J. Inst. Met.* **92**, 292.
- CWILONG, B. M. (1945) *Nature (London)* **155**, 361.
- CZOCHRALSKI, T. (1913) *J. Phys. Chem.* **92**, 239.
- DARWIN, C. G. (1922) *Philos. Mag.* **43**, 800.
- DASH, W. C. (1958) *J. Appl. Phys.* **29**, 736.
- DAVIES, V. DE L. and WEST, J. M. (1963–4) *J. Inst. Met.* **92**, 175.
- DAY, M. G. (1970) *J. Inst. Met.* **98**, 57.
- DAY, M. G. and HELLAWELL, A. (1967) *J. Inst. Met.* **95**, 377; (1968) *Proc. R. Soc. London, Ser. A* **305**, 473.
- DEAN, F. H., KERR, J. M., and HELLAWELL, A. (1962) *J. Inst. Met.* **90**, 253.
- DELVES, R. T. (1965) *J. Inst. Met.* **90**, 254; (1968) *Br. J. Appl. Phys.* **16**, 345.
- DOHERTY, R. and DAVIS, R. S. (1959) *Acta Metall.* **7**, 118.
- DOHERTY, R., CANTOR, B., and FAIRS, S. (1973) *Metall. Trans.* **90**, 127.
- DOUBLE, D. and HELLAWELL, A. (1969) *Acta Metall.* **17**, 1071; (1971) *Ibid.* **19**, 1303.
- DUWIZ, P. (1950) *Trans. Am. Soc. Met.* **60**, 605.
- ELBAUM, C. and CHALMERS, B. (1955) *Can. J. Phys.* **33**, 196.
- ELLWOOD, E. C. and BAGLEY, K. A. (1949) *J. Inst. Met.* **76**, 631.
- ESAKA, H. and KURZ, W. (1984) *J. Cryst. Growth* **69**, 362.
- EWING, H. H. (1971) *J. Cryst. Growth* **11**, 221.
- FEHLING, J. and SCHEIL, E. (1962) *Z. Metallkd.* **56**, 383.
- FLEMINGS, M. C. (1974) *Solidification Processing*, McGraw-Hill, New York.
- FLEMINGS, M. C., RIEK, R. G., and YOUNG, K. P. (1976) *Int. Casting Met. J.* **1**, 11.
- FRANK, F. C. (1952) *Adv. Phys.* **1**, 91; (1956) *Deformation and Flow of Solids (I.U.T.A.M. Colloquium)*, p. 73, Springer-Verlag, Berlin; (1956–7) *J. Inst. Met.* **85**, 581; (1958) *Growth and Perfection of Crystals*, p. 306, Wiley, New York.
- GLICKSMAN, M. and CHILDS, V. A. (1962) *Acta Metall.* **10**, 925.
- GLICKSMAN, M. E. and SCHAEFER, R. J. (1967) *J. Cryst. Growth* **1**, 237; (1968a) *Ibid.* **2**, 219; (1968b) *Solidification of Metals*, p. 43, Iron and Steel Institute, London.
- GLICKSMAN, M. E. and VOLD, C. M. (1969) *Ser. Metall.* **5**, 493.
- GOMSALL, S. Y., SHIRAIISHI, S. Y., and WARD, R. G. (1966) *J. Aust. Inst. Met.* **10**, 220.
- GOSWAMI, R., CHATTOPADYAY, K., KIM, W. T., and CANTOR, B. (1992) *Metall. Trans.* **23A**, 3207.
- GRAF, L. (1951) *Z. Metallkd.* **42**, 336.
- GULLIVER, C. H. (1913) *J. Inst. Met.* **9**, 11.
- HELLAWELL, A. (1962) *Personal communication*; (1973) *Prog. Mater. Sci.* **15**, 1.
- HELLAWELL, A. and HERBERT, P. M. (1962) *Proc. R. Soc. London, Ser. A* **269**, 560.
- HENSEL, F. (1937) *Trans. Metall. Soc. AIME* **124**, 300.
- HOLLOMON, J. H. and TURNBULL, D. (1951) *Trans. Metall. Soc. AIME* **191**, 3; (1953) *Prog. Met. Phys.* **4**, 333.
- HORVAY, G. A. and CAHN, J. W. (1961) *Acta Metall.* **9**, 695.
- HUANG, S. C. and GLICKSMAN, M. E. (1981) *Acta Metall.* **29**, 701.
- HUNT, J. D. (1979) *Solidification and Casting of Metals*, The Metals Soc., London, p. 3; (1990) *Acta Metall. Mater.* **38**, 411; (1991) *Ibid.* **39**, 2117; (1997) *Mater. Sci. Technol.* **18**, 114.
- HUNT, J. D. and JACKSON, K. A. (1966) *Trans. Metall. Soc. AIME* **236**, 843.
- HUNT, J. D. and LU, S.-Z. (1996) *Met. and Mater. Trans.* **27A**, 611; (1991) *Mater. Trans.* **27**, 474.
- HUNT, J. D., SPITTLE, J. A., and SMITH, R. W. (1968) *Iron Steel Inst., London, Publ.* **110**, 57.
- HUNTER, G. A. and CHADWICK, G. A. (1972) *J. Iron Steel Inst.* **210**, 117.
- IVANTSOV, G. P. (1947) *Dokl. Akad. Nauk. SSSR* **58**, 567.
- JACKSON, K. A. (1958a) *Liquid Metals and Solidification*, p. 178, American Society for Metals, Metals Park, OH; (1958b) *Growth and Perfection of Crystals*, p. 319, Wiley, New York.

- JACKSON, K. A. and HUNT, J. D. (1965) *Acta Metall.* **13**, 1212.
- JAFFEY, H. and CHADWICK, G. A. (1968) *Philos. Mag.* **18**, 573.
- JONES, D. R. H. and CHADWICK, G. A. (1970) *Philos. Mag.* **28**, 291.
- KATTAMIS, T. Z. and FLEMINGS, M. C. (1966) *Trans. Metall. Soc. AIME* **236**, 1523.
- KESSLER, D. and LEVINE, H. (1986) *Phys. Rev. Lett.* **59**, 3669.
- KHAIKIN, S. E. and BENET, H. P. (1939) *Dokl. Akad. Nauk. SSSR* **23**, 31.
- KIM, W. T. and CANTOR, B. (1991) *J. Mater. Sci.* **26**, 2628; (1992) *Acta Metall. Mater.* **40**, 3339; (1994) *Ibid.* **42**, 3115.
- KOTLER, G. R. and TARSHIS, L. A. (1968) *J. Cryst. Growth* **3**, 4; (1969) *Ibid.* **5**, 92.
- KRAMER, T. J., MASSION, G. F., and TILLER, W. A. (1963) *Trans. Metall. Soc. AIME* **227**, 3347.
- LANGER, J. S. and MÜLLER-KRUMBHAR, H. (1977) *J. Cryst. Growth* **42**, 11; (1978a) *Acta Metall.* **24**, 1681; (1978b) *Ibid.* **24**, 1689; (1978c) *Ibid.* **24**, 1697.
- LIPTON, J., KURZ, W., and TRIVEDI, R. (1987) *Acta Metall.* **35**, 957.
- LIVINGSTON, J. D. and CLINE, H. E. (1969) *Trans. Metall. Soc. AIME* **245**, 351.
- LOTHE, J. and POUND, G. M. (1962) *J. Chem. Phys.* **36**, 208.
- LU, S.-Z. and HUNT, J. D. (1992) *J. Cryst. Growth* **123**, 17.
- LUX, B., KURZ, W., and GRAGES, M. (1969) *Prakt. Metallogr.* **6**, 464.
- MEIRON, D. (1986) *Phys. Rev. A* **33**, 2704.
- MENDENHALL, C. E. and INGERSOLL, L. R. (1908) *Philos. Mag.* **15**, 205.
- METALLURGICAL SOCIETY OF AIME (1961) *Physical Chemistry of Process Metallurgy*, ed. G. R. St. Pierre, p. 845, Interscience, New York.
- MILLER, W. R. and CHADWICK, G. A. (1967) *Acta Metall.* **15**, 607.
- MIYAZAWA, Y. and POUND, G. M. (1976) *J. Cryst. Growth* **23**, 45.
- MOORE, K. I., ZHANG, D. L., and CANTOR, B. (1990) *Acta Metall. Mater.* **38**, 1327.
- MORRIS, L. R. and WINEGARD, W. C. (1969a) *J. Cryst. Growth* **6**, 61; (1969b) *J. Inst. Met.* **97**, 220.
- MORROGH, H. and WILLIAMS, W. D. (1968) *J. Iron Steel Inst.* **89**, 305.
- MULLINS, W. W. and SEKERKA, R. F. (1963) *J. Appl. Phys.* **34**, 323; (1964) *Ibid.* **35**, 444.
- NASH, C. E. and GLICKSMAN, M. E. (1974a) *Acta Metall.* **22**, 1283; (1974b) *Ibid.* **22**, 1291.
- NORTHCOTT, L. (1938) *J. Inst. Met.* **62**, 101; (1939) *Ibid.* **65**, 173; (1946) *Ibid.* **72**, 283.
- PAPAPIETROU, A. (1935) *Z. Kristallogr., Kristallgeom., Kristallphys., Kristallchem.* **92**, 108.
- PEREPEZKO, J. H. (1980) *Rapid Solidification Processing: Principles and Technology*, eds. R. Mehrabian, B.H. Kear and M. Cohen, p. 36, Caiteurs, Baton Rouge.
- PEREPEZKO, J.H. and PAIK, J. S. (1984) *J. NonCryst. Solids*, **61/62**: 113.
- PEREPEZKO, J. H., RASMUSSEN, B. J., ANDERSON, I. E., and LOPER, C. R. (1979) *Solidification and Casting of Metals*, p. 169, The Royal Society of London.
- PEANN, W. G. (1952) *Trans. Metall. Soc. AIME* **194**, 747; (1957) *Solid State Phys.* **4**, 423; (1958) *Zone Melting*, 2nd edn., Kruger, New York.
- PLASKETT, T. G. and WINEGARD, W. C. (1960) *Can. J. Phys.* **38**, 1077.
- POUND, G. M. and LA MER, V. K. (1952) *J. Am. Chem. Soc.* **74**, 2323.
- POWELL, E. G. (1965) *J. Aust. Inst. Met.* **10**, 223.
- PUMPHREY, W. I. and LYONS, J. V. (1951) *Nature (London)* **163**, 960.
- RASMUSSEN, D. H. and LOPER, C. R. (1975) *Acta Metall.* **23**, 1715.
- READ, W. T. (1951) *Dislocations in Crystals*, McGraw-Hill, New York.
- ROSENBERG, A. and TILLER, W. A. (1957) *Acta Metall.* **5**, 565.
- RUTTER, J. W. and CHALMERS, B. (1953) *Can. J. Phys.* **31**, 15.
- SAITO, F., GOLDBERG, H., and MÜLLER-KRUMBHAR, H. (1988) *Phys. Rev. Lett. A* **35**, 2166.
- SAKA, H., NISHIKAWA, K., and IMURA, T. (1987) *Philos. Mag. Lett.* **55**, 163.
- SCHAEFFER, P. C. R. J. (1948) *Bull. Am. Meteorol. Soc.* **29**, 175.
- SCHEEL, E. (1942) *Z. Metallkd.* **34**, 70; (1946) *Ibid.* **37**, 1.
- SCHOECK, C. and TILLER, W. A. (1960) *Philos. Mag.* **5**, 43.
- SCRIPOU, V. (1977) *Crystal Growth and Materials*, eds. E. Kaldis and H. Scheel, p. 1327. North Holland, Amsterdam.
- SEKERKA, R. F. (1965) *J. Appl. Phys.* **36**, 264; (1967) *J. Phys. Chem. Solids* **28**, 983.
- SHARP, R. S. and HELLAWELL, A. (1971) *J. Cryst. Growth* **8**, 29.
- SHENG, H. W., XU, J., YU, L. G., SUN, X. K., HU, Z. Q., and LU, K. (1996) *J. Mater. Res.* **11**, 2841.

- SHEWMON, P. (1957) *Acta Metall.* **5**, 335.
- SHIRAIISHI, S. Y. and WARD, R. G. (1964) *Con. Met. Ind.* **3**, 117.
- SMIALOWSKI, M. (1937) *Z. Metallkd.* **29**, 133.
- SMITH, V. G., TILLER, W. A., and RUTTER, J. W. (1955) *Can. J. Phys.* **33**, 723.
- SOUTHIN, R. T. and CHADWICK, G. A. (1978) *Acta Metall.* **26**, 223.
- SPENCER, D. B., MEHRABIAN, R., and FLEMINGS, M. C. (1972) *Metall. Trans.* **3**, 1925.
- STOWELL, M. J. (1970) *Phil. Mag.* **22**, 1.
- STOWELL, M. J., LAW, T. T., and SMART, J. (1970) *Proc. R. Soc. London, Ser. A* **318**, 231.
- STRAUMANIS, M. and BRAKSS, N. (1935) *Z. Phys. Chem.* **B30**, 117.
- SWORN, C. H. and BROWN, T. E. (1972) *J. Cryst. Growth* **12**, 195.
- TAKAGI, M. (1954) *J. Phys. Soc. Jpn.* **9**, 359.
- TAKAHASHI, T. and TILLER, W. A. (1969) *Crystal Growth*, ed. H.S. Peiser, p. 709, Pergamon Press, London, 1967.
- TANNER, B. K. (1972) *J. Cryst. Growth* **16**, 8; (1973) *Z. Naturforsch.* **28**, 5.
- TARSHIS, L. A., WALKER, R. J., and RUTTER, J. W. (1971) *Metall. Trans.* **2A**, 2875; (1950a) *J. Appl. Phys.* **21**, 1022; (1950b) *J. Chem. Phys.* **96**, 3; (1963) *Acta Metall.* **1**, 453.
- TEGHTSOONIAN, E. and CHALMERS, B. (1951) *Can. J. Phys.* **29**, 370; (1952) *Ibid.* **30**, 388.
- TIEMKIN, D. E. (1960) *Dokl. Akad. Nauk. SSSR* **132**, 1307.
- THALL, B. M. and CHALMERS, B. (1949) *J. Inst. Met.* **78**, 79.
- THOMAS, D. G. and STAVELEY, L. A. K. (1952) *J. Chem. Soc.* 4569.
- TILLER, W. A. (1958) *Liquid Metals and Solidification*, p. 276, American Society for Metals, Cleveland, OH.
- TILLER, W. A., JACKSON, K. A., RUTTER, J. W., and CHALMERS, B. (1953) *Acta Metall.* **1**, 128.
- TRIVEDI, R. (1970) *Acta Metall.* **18**, 287.
- TRIVEDI, R. and KURZ, W. (1994a) *Acta Metall. Mater.* **42**, 13; (1994b) *Int. Mater. Rev.* **39**, 49.
- TRIVEDI, R. and MASON, J. T. (1991) *Metall. Trans.* **22A**, 215.
- TURNBULL, D. (1950a) *J. Metals*, **188**, 1144; *J. Appl. Phys.* **21**, 1022; (1950b) *J. Chem. Phys.* **18**, 68; (1952a) *Ibid.* **20**, 411; TURNBULL, D. (1950b) *J. Chem. Phys.* **18**, 768. (1952b) *Ibid.* **20**, 1327.
- TURNBULL, D. and CECIL, R. E. (1950) *J. Appl. Phys.* **21**, 804.
- TURNBULL, D. and CECIL, R. E. (1950c) *Trans. Amer. Inst. Min. (Metall.) Engrs.*, **206**, 124.
- TURNBULL, D. and CORMIA, R. L. (1961) *J. Chem. Phys.* **34**, 820.
- TURNBULL, D. and VONNEGUT, B. (1952) *Ind. Eng. Chem.* **46**, 1292.
- UHMANN, D. and CHADWICK, G. A. (1961) *Acta Metall.* **9**, 235.
- VONNEGUT, B. (1948) *J. Colloid Sci.* **3**, 563.
- WALTON, D. and CHALMERS, B. (1959) *Trans. Metall. Soc. AIME* **215**, 417.
- WALTON, D., TILLER, W. A., RUTTER, J. W., and WINEGARD, W. C. (1955) *Trans. Metall. Soc. AIME* **203**, 1023.
- WAN, X., HAN, Q., and HUNT, J. D. (1997) *Acta Metall. Mater.* **45**, 3975.
- WANG, C. C. and SMITH, C. S. (1950) *Trans. Metall. Soc. AIME* **188**.
- WASHBURN, J. and NADEAU, J. (1958) *Acta Metall.* **6**, 665.
- WEINBERG, F. and CHALMERS, B. (1951) *Can. J. Phys.* **29**, 382; (1952) *Ibid.* **30**, 488.
- WEINBERG, F. and TEGHTSOONIAN, E. (1957) *Acta Metall.* **5**, 455.
- ZHANG, D. L. and CANTOR, B. (1990a) *Mater. Sci. Eng. A* **121**, 209; (1990b) *Philos. Mag.* **62A**, 557; (1990c) *J. Cryst. Growth* **104**, 511; (1992) *Acta Metall. Mater.* **40**, 2951.
- ZHANG, D. L., HUTCHISON, J. L., and CANTOR, B. (1994) *J. Mater. Sci.* **29**, 2147.
- ZIEF, M. and WILCOX, W. R. (1967) *Fractional Solidification*, Marcel Dekker, Amsterdam.

## CHAPTER 15

# *Polymorphic Changes*

### 70. SOLID-STATE REACTIONS IN PURE METALS

The change from one equilibrium crystal structure to another structure of the same composition is the simplest type of transformation in the solid state. Such reactions occur frequently in the metallic elements (some 20 of which exhibit allotropy or polymorphism), in intermetallic compounds and in some intermediate solid solutions. A thermodynamic discussion of the polymorphic changes in elements was given in Section 16. A polymorphic change may be either a nucleation and growth process or a martensitic transformation, and some changes may exhibit either or both mechanisms depending on the imposed conditions of reaction. In this chapter, attention is confined to changes of the nucleation and growth type.

It is not uncommon for structural transformation in a pure metal to be effected by a nucleation and growth reaction at temperatures near to the equilibrium transformation temperature, but by a martensitic reaction at temperatures sufficiently far removed from the equilibrium temperature. This statement is based partly on experiments in which high speed thermal analysis techniques are used in conjunction with rapid gas quenching (Greninger, 1942; Duwez, 1951). If the cooling rate from the high temperature phase is progressively increased, the thermal arrest sometimes occurs at progressively lower temperatures, but reaches a limiting temperature at fast cooling rates. Quenching is thus able to achieve only a limited suppression of the transformation in such metals, and examination of originally flat surfaces after quenching above the critical rate shows the characteristic distortions associated with a martensitic change. Direct evidence of the transition from a nucleation and growth reaction to a martensitic reaction was obtained by Gaunt and Christian (1959) by hot stage microscopy for zirconium and by Bibby and Parr (1964) for iron. It should be remarked here, however, that the thermal analysis results for iron are more complex than is indicated by the above simple description, and several different plateaux are observed as the cooling rate is increased (see below). However, the transitions from  $\beta$  to  $\alpha$  manganese can apparently be produced only by nucleation and growth type mechanisms, and sufficiently fast quenching can result in the retention of the high temperature form at low temperatures for indefinite periods of time. There are also some transformations (in cobalt, lithium and sodium for example) in which the equilibrium transition temperature is so low that only martensitic reaction is feasible,

there being insufficient thermal energy for thermally activated growth. As would be expected, the speed of a polymorphic change is much lower than that of a liquid transition, but is appreciably faster than that of a solid-state phase reaction in which changes in compositions are produced. Heating and cooling curves, taken at rates attainable in normal laboratory apparatus, show that both superheating and supercooling of a particular transition may be appreciable but, with the exceptions noted above, it is rarely possible to quench a pure metal sufficiently rapidly to retain the metastable high temperature structure indefinitely. However, in some cases, addition of a very small amount of an alloying element to a metal will slow the reaction sufficiently to allow the high temperature form to be retained in the quench. If the transition is from one disordered solid solution to another of the same composition, it is not obvious why a solute should have much influence on the growth; however, if the solute atoms have to be displaced relative to the solvent, an effect on the overall velocity of reaction is understandable. The effects of iron and copper on the  $\beta$ - $\alpha$  transformation in manganese are interesting. Iron, which is soluble in both forms of manganese, slows the reaction, but copper, which is insoluble in both forms, increases the rate of reaction considerably. Presumably, the precipitated copper crystals (precipitated from solution in  $\gamma$ -manganese) act as effective nucleation sites for the change.

As mentioned above, data plotted as arrest temperature versus cooling rate in the case of iron indicate the existence of one or more plateaux at temperatures higher than  $M_s$  (see, e.g., Wilson, 1970; Morozov *et al.*, 1971, 1972). This has been attributed to the production of various non-martensitic structures with different morphologies; for example, a product with a crystallographic interface might be nucleated more readily than a product with a more mobile, incoherent interface (Hawbolt and Massalski, 1971). Wilson (1970, 1994) has cited experimental evidence in support of the view that a plateau temperature obtained in a series of continuous cooling experiments coincides with the temperature of the maximum rate of transformation under isothermal conditions, i.e. with the "nose" of a C-shaped  $T$ - $T$ - $T$  curve. There is some metallographic evidence to support the view that the nature of the product changes at each plateau temperature, but Bhattacharyya *et al.* (1973a) suggested that the plateaux are artefacts resulting from the method of plotting the data and that they disappear if the arrest temperatures are plotted against the logarithms of the cooling rate, as was done by Duwez (1951), or against the logarithm of the time taken to cool from the thermodynamic equilibrium temperature ( $T_0$ ) to the arrest temperature. However, the data of Wilson and of the Russian workers include several plateaux which cannot be made to disappear merely by plotting against the logarithm of the cooling rate.

Measurements of arrest temperature and growth rate strongly suggest that the growth rate is rapid in comparison with the nucleation rate. An analysis of non-isothermal transformation (Bhattacharyya *et al.*, 1973b) is based on the hypothesis that, for supercoolings of up to 100 °C in a typical case, the time taken to form a nucleus dominates the kinetics and so governs the arrest temperature at a typical cooling rate. The variation of arrest temperature with cooling rate is then assumed to be determined by the variation with temperature of the critical nucleus size and of the rate of growth of subcritical

embryos. This does not necessarily apply to transformations at temperatures where the atomic mobility is limited and growth rates are correspondingly slow.

The polymorphic transition in pure tin is the most familiar example of a phase change in a pure material and was known to Aristotle and Plutarch. The high temperature ( $\beta'$ ) form has a tetragonal structure and is called white tin. The low temperature ( $\alpha$ ) form, stable below about 20°C, has the diamond cubic structure and is called grey tin. Under normal circumstances, transformation to grey tin does not occur until the sample has been cooled very appreciably below 20°C. There is a very large volume increase during the  $\beta$ - $\alpha$  change, and samples of grey tin are commonly obtained as a polycrystalline powder or as severely cracked and crumbling aggregates. Special techniques are required to produce single crystals of grey tin.

The large volume change is undoubtedly the major factor inhibiting transformation. It was pointed out in Section 52 that the strain energy resulting from the formation of a nucleus constrained by a surrounding matrix can be very high. This is particularly true of the tin transition, where the strain energy to form a spherical crystal of grey tin is about 10 times the free energy of transformation at 0°C. The volume change has two consequences. The first of these is that homogeneous nucleation of the  $\alpha$  phase from the  $\beta$  phase is impossible and, even in imperfect specimens, the nucleation rate will be low. This conclusion is fully supported by the large amount of available data. Experience shows that the nucleation rate is very sensitive to cold working, impurity content, particle irradiation, surface condition and previous history. The particularly favourable experimental conditions allow the importance of nucleation to be demonstrated directly, inasmuch as the transformation can be initiated by rubbing the surface of the white tin with small particles of grey tin ("inoculation"). This is the solid-state equivalent of the "seeding" process often used to prevent supercooling during freezing, or to promote crystallization from solution.

In some investigations, nuclei of grey tin seemed to form spontaneously, even in samples of white tin which were far from perfect. Becker (1958) reported the absence of spontaneous transformation in samples after 12 months at -30°C or 6 months at -79°C, and nucleation could not be induced either by deformation or by electrolytic dissolution of the surface. Spontaneous transformation was observed in other, apparently equivalent, samples after some grey tin had been brought into the laboratory, suggesting that accidental inoculation had taken place, although special precautions were taken to prevent this. Whether or not inoculation is necessary, it seems probable that nucleation always occurs at the free surface. This would explain the considerable influence of oxide films and immersion in suitable electrolytes (see, e.g., Smith and Raynor, 1957). The second effect of the large volume change is that very large stresses are produced in both matrix and product as a region of grey tin grows. The matrix, being ductile, accommodates these to some extent by deformation, but the stresses in the product soon cause cracking and lead to the characteristic disintegrated form. Compact pieces of grey tin have been obtained by transforming thin slices of white tin (Groen and Burgers, 1954) and wires (Ewald and Kohnke, 1955). The experiments of Groen and Burgers suggest that the necessary conditions for obtaining compact pieces of grey tin are the use of thin sheets, a slow rate of

transformation, and the growth of only one nucleus produced by inoculation. Groen and Burgers found it possible to obtain single crystals of grey tin from crystals of white tin if the tin contained  $\sim 0.1$  at.% mercury.

The investigations of Burgers and his collaborators on the mechanism and kinetics of both the  $\beta$ - $\alpha$  transformation and its inverse have been summarized by Burgers and Groen (1957). In this work, and that of Hall (1956), no evidence of orientation relations between parent and product phases was found, as might be anticipated from the observed lack of coherence. The cooling transformation shows all the usual features of a nucleation and growth reaction, and the product regions are approximately spherical. Burgers and Groen point out that, during the heating transformation, small domains of white tin appear suddenly and grow to their final size in 7–30 s. A small number of such domains form in each individual grain of grey tin, sometimes at intervals of several minutes. The volume change requires the white tin particles to shrink, and this leads to cracking and disintegration of the grey tin matrix. The final size of each white tin particle is approximately constant, possibly because growth is stopped by cracking. Large single crystals of grey tin, produced by growth from a tin–mercury alloy, were found to transform by the stepwise displacement of a planar interface (Ewald and Tufte, 1958), thus suggesting some interface coherence. The linear growth rate of white tin from grey tin is shown in Fig. 15.1 as a function of temperature; the strong variation indicates that the growth is thermally activated. The growth rate in Fig. 15.1 is much larger than the corresponding growth rate measured for the inverse transformation on cooling.

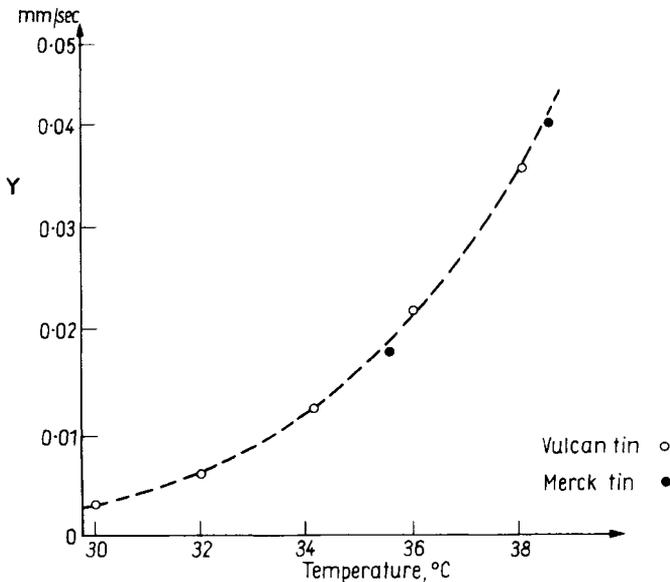


FIG. 15.1. The temperature dependence of the linear growth rate of white ( $\beta$ ) tin from grey ( $\alpha'$ ) tin (after Burgers and Groen, 1957).

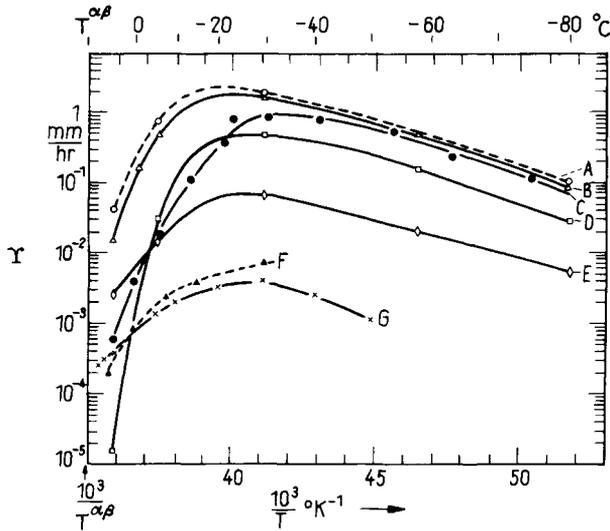


FIG. 15.2. The temperature dependence of the linear growth rate of grey ( $\alpha$ ) tin from white ( $\beta$ ) tin for samples of differing impurity content (after Becker, 1958). Samples *B*, *C* and *F* are different "high purity" samples. Samples *A*, *D* and *E* are sample *C* + 0.002 at.% Al, 1.0 at.% Au and 1.0 at.% Cu respectively. Curve *G* represents old data of Tamman and Dreyer for Kahlbaum tin.

Figure 15.2 shows some of Becker's results for the growth rate of grey tin from white tin, and illustrates the considerable effects of impurities. The general form of Fig. 15.2, with a maximum growth rate at a particular temperature, has been verified by many other workers, although the temperature of the maximum seems to vary somewhat with the purity and initial condition of the tin. The form of the curve may be described in terms of chemical reaction rate theory, as previously discussed in Section 53. However, it is not easy to obtain quantitative agreement with a particular molecular mechanism.

Cagle and Eyring (1953) applied reaction rate theory to some earlier results of Tamman and Dreyer (also shown in Fig. 15.2), and they obtained approximate agreement by making the assumption that only about  $10^{-10}$ – $10^{-13}$  of the interface atoms are at reactive sites. A more promising approach appears to be that of Becker (1958), who applied the analysis outlined in Section 53. Figure 15.3 shows his plot of the function  $\ln X$  against  $1/T$  [see eqn. (53.10)]. This plot tends to a straight line at low temperatures and, assuming  $(T^2/x_0)(\partial x_0/\partial T)$  to be negligible at these temperatures, Becker obtained a value for the heat of activation. The theoretical curve of  $X$  against  $1/T$  which is then obtained from this value is adjusted so that the maximum rate agrees with the experimental value, and this requires  $x_0 = 10^{-2}$ – $10^{-4}$ . The theoretical curve, however, shows its maximum rate at a higher temperature than that observed, and this is attributed to the assumption of constant  $x_0$ . Figure 15.3 shows, in fact, that  $(\partial x_0/\partial T)$  is not zero at all temperatures and the experimental results can be explained if  $(1/x_0)(\partial x_0/\partial T)$  increases with increasing temperature. Direct measurement of the temperature dependence of  $x_0$  to try to justify

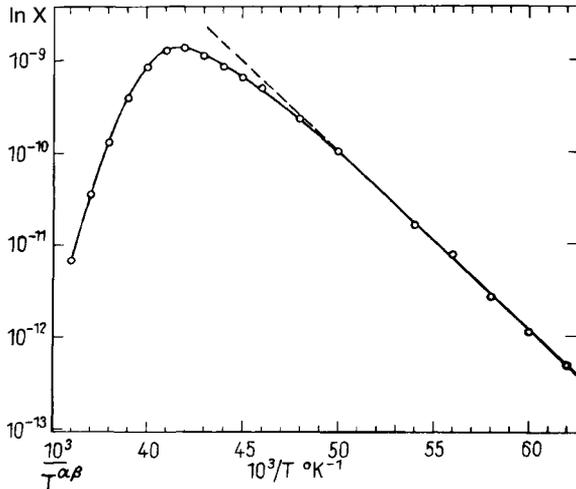


FIG. 15.3. Plot of the function  $\ln X$  (see text) versus  $1/T$  for the growth of grey tin (after Becker, 1958).

this assumption does not appear to have been attempted but Becker's theory does explain both the marked dependence of the growth rate on the impurity content (caused by change in  $\Delta_i h^*$ ) and the relative insensitivity of the temperature of maximum growth rate to impurities.

Observations on the kinetics of the  $\beta$ - $\alpha$  transformation show a marked dependence on the form of the specimen and its previous history. In the work of Burgers and Groen, long induction periods for the nucleation of  $\alpha$  regions were necessary for previously untransformed specimens but, after several transformation cycles, a small number of  $\alpha$  nuclei formed in each  $\beta$  particle at the beginning of transformation. This nucleation apparently results from regions of grey tin which had been incompletely transformed in the previous half cycle. The authors found evidence that neither the heating nor the cooling transformations were complete under normal conditions. Plots of  $\log \log(1/1 - \zeta)$  against  $\log t$  are linear with a slope of three for specimens in which nuclei are present at the onset of transformation, in agreement with Table IX on p. 546 for zero nucleation rate and three-dimensional growth. A value for the growth rate may be deduced from the time to half transformation, and results obtained in this way are in good agreement with the direct measurements.

As mentioned above, there is also an induction period for nucleation of the  $\alpha$  to  $\beta$  change. Burgers and Groen pointed out that, under the usual conditions of this reaction, the transformation rate is determined almost entirely by the nucleation rate. For constant nucleation rate, the kinetics will be described by

$$\zeta = 1 - \exp(-vIvt) \quad (70.1)$$

if each nucleus is assumed to grow to a fixed volume  $v$ . This is equivalent to the Avrami equation with  $n = 1$ . Experimental plots of  $\log \log(1/1 - \zeta)$  against  $\log t$  show straight lines

but with slopes varying from one to 2.4. The larger values can be explained if there is a transient in the nucleation rate which is then increasing with time.

An interpretation of the induction period for nucleation in terms of the theory of transient nucleation described in Section 50 has been given by Dunning (1957). Using eqn. (50.7) he shows that eqn. (50.8) may be written in the form

$$\tau = (C/\Delta T^-)^2 \exp(\Delta_a h^*/kT) \quad (70.2)$$

where  $\Delta T^-$  is the superheating and  $\Delta_a h^*$  is the free energy of activation crossing the interface. Assuming  $\tau$  to be proportional to the observed nucleation period, a value of  $\Delta_a h^*$  may be estimated as 13.1 kcal g atom<sup>-1</sup> (as compared with the value of 14 obtained by plotting the results in Fig. 15.1 as  $\ln T$  versus  $1/T$ ). The agreement supports the interpretation as the same thermally activated process probably controls both the nucleation and later stages of the transformation. This follows because the interface is probably incoherent even when the nucleus is quite small.

Isothermal transformation curves have also been determined for the  $\beta$ - $\alpha$  change in pure manganese (Hubbard *et al.*, 1959). The  $T$ - $T$ - $T$  diagram shows a maximum transformation rate at about 600°C, some 100°C below the thermodynamic transformation temperature. The transition in both directions is extremely sluggish, and the arrests on heating and cooling curves show considerably hysteresis. Nevertheless, very fast cooling rates have to be used to suppress the  $\beta$ - $\alpha$  transformation completely in pure manganese, although it is easy to do this with a slightly impure manganese.

Some typical  $\log \log(1/1 - \zeta)$  versus  $\log t$  curves were shown in Fig. 12.1; those above the nose of the  $T$ - $T$ - $T$  curves were determined by electrical resistance measurements and those at lower temperature by measurements of integrated X-ray intensities. The slopes of the lines at most temperatures suggest random nucleation and three-dimensional growth, but direct confirmation of this was not possible because of the difficulty of microscopic observation of the very brittle manganese. The volume change in the transformation is comparatively small, and the difficulties in nucleation implied by the suppression are probably due entirely to the complex structures and large unit cells of the two forms of manganese. It seems unlikely that the two structures can fit together to give a coherent interface so the surface energy presumably includes a large misfit term, leading to a low nucleation rate. Little is known about imperfections in complex structures like  $\alpha$  and  $\beta$  manganese, so it is uncertain whether or not nucleation always begins from grain boundaries.

Interesting evidence of the importance of imperfections for nucleation of a polymorphic change comes from work on the precipitation of iron particles from solid solution in copper. In agreement with earlier suggestions, Newkirk (1957) established that the iron precipitates from the copper as discrete particles both of coherent  $\gamma$  phase and of  $\alpha$  iron, the former predominating in the initial stages. The  $\gamma$  iron, which is thermodynamically unstable, can be induced to redissolve and precipitate on the  $\alpha$  particles, but the interesting observation with regard to nucleation is that the  $\gamma$  particles do not transform spontaneously to  $\alpha$  even on cooling to  $-196^\circ\text{C}$ . This is in marked contrast to the

behaviour of bulk iron which cannot be supercooled more than a few degrees below the equilibrium transformation temperature of  $910^{\circ}\text{C}$ . The small coherent particles are presumably sufficiently perfect to prevent the formation of nuclei, although they transform at once if the specimen is plastically deformed. In a similar manner, it has been observed that highly perfect iron "whiskers" do not transform from the  $\alpha$  to the  $\gamma$  phase until a superheating of about  $200^{\circ}\text{C}$  has been achieved.

## 71. MASSIVE TRANSFORMATIONS

Some solid solutions transform very rapidly to a new single-phase structure with the same composition as the original phase. When a reaction of this type exhibits nucleation and growth rather than martensitic characteristics, it is now usually called a massive transformation. The name is derived from the microstructural appearance and appears to have been first used by Greninger (1939) in connection with the change which occurs in certain copper-aluminium alloys. Some of the best-known massive transformations occur in critical composition ranges of alloys of copper with zinc, aluminium, germanium, etc., and the product in these alloys may then be regarded as a low temperature supersaturated phase in the equilibrium diagram. Massive transformations may also be obtained in iron-rich alloys on rapid cooling from the  $\gamma$  phase.

The product of a massive transformation may be either an equilibrium phase or a supersaturated solid solution. The thermodynamic condition for an alloy of any composition to undergo a massive transformation from the  $\beta$  to the  $\alpha$  phase is simply that the free energy shall decrease; i.e. the specimen must be cooled below the  $T_0$  temperature (at which the two phases have the same free energy appropriate to the composition of the alloy). This means that the massive transformation can occur within a two-phase region of the equilibrium diagram, in which case the product will be a supersaturated  $\alpha$  solid solution, or within a single-phase region, in which case the product will be the equilibrium  $\alpha$  phase.

The kinetic features of a massive transformation are superficially similar in many ways to those of a martensitic transformation, and there are some alloys which transform in both ways. The similarities arise from the high speed of both kinds of transformation, which are often found to have taken place during rapid cooling from the high temperature phase. Nevertheless, the mechanism of the massive change is quite distinct from that of a martensitic transformation, and the use of terms like "massive martensite" is not to be encouraged as it confuses the situation. The crystals of a massive phase usually have a jagged, rather irregular appearance with a number of short planar segments and no particular overall shape. Despite the planar segments, there is no evidence of the tilting of a previously polished surface as in a martensitic transformation, and it is generally believed that the interphase boundary is incoherent. Massalski (1958) showed by metallographic and X-ray techniques that there is no systematic crystallographic relationship between the crystals of the two phases, and the massive crystals frequently cross pre-existing grain boundaries of the parent phase with apparently little difficulty.

The growth mechanism of a massive transformation thus requires the thermally activated migration of individual atoms through small distances and in particular across the incoherent interface. This mechanism does not differ significantly from that of a non-martensitic polymorphic transition in a pure metal.

No quantitative measurements of growth rate appear to have been recorded but, as complete transformation is often achieved at rapid cooling rates, it seems likely that the activation energy for growth is small, and is probably comparable to that for grain boundary rather than lattice diffusion. However, nucleation of the incoherent interface may not be easy, and the supercooling needed to achieve the transformation is intermediate between that giving reaction at high temperatures and the martensitic transformation at low temperatures. Nucleation of the martensite is more difficult because of the large strain energy, but its growth, not requiring thermal energy, is even more rapid.

Many experimental investigations have been made of the temperature of the first thermal arrest during a cooling curve as a function of the cooling rate from the high temperature phase. It follows from the above discussion that if all three types of reaction are possible, the thermal arrest which signals the initiation of transformation will occur at progressively lower temperatures as the cooling rate is increased from a low value. However, at some critical cooling rate, the temperature at which the precipitation reaction begins will have been lowered to that at which the massive transformation is nucleated, and the transformation will then proceed massively much more rapidly. For high cooling rates, the arrest temperature is apparently unchanged, leading to a plateau in the plot of this temperature against cooling rate. Wilson (1970, 1984) suggested that the temperature of this plateau will correspond to that of the maximum isothermal rate of transformation. Eventually, however, the time available at the plateau temperature will become inadequate for nucleation and growth of the massive product and an increased driving force will be necessary, leading to a further decrease in arrest temperature. This will continue until the temperature for martensitic nucleation ( $M_s$ ) is reached without prior transformation when a second plateau will be produced. Thus from this point of view, a martensitic transformation will never be achieved unless the cooling rate is sufficiently large to suppress the potential massive transformation. In practice, this condition may either be too stringent to be feasible or may be achieved at almost any cooling rate.

Much attention has been devoted to the transformations in iron and its alloys during continuous cooling from austenite. Gilbert and Owen (1962) plotted the arrest temperature against cooling rate for Fe-Ni, Fe-Cr and Fe-Si alloys and found for each alloy composition a plateau region (see Fig. 15.4) corresponding to the formation of ferrite by a massive transformation.

Later work has shown that the behaviour of iron alloys is very much more complex than is indicated by the simple description above. Thus Mirzayev *et al.* (1987) found four plateaux (Fig. 15.5) and Wilson (1970, 1984, 1994) found five (Fig. 15.6). The two lowest plateaux in both of these investigations are ascribed to two different forms of martensite ("lath" or "twinned") and the one above this represents bainitic ferrite which is also believed to form by a shear mechanism (see Chapter 25). The first plateau, or the first two according to Wilson, corresponds to genuine massive transformation. Wilson calls the first

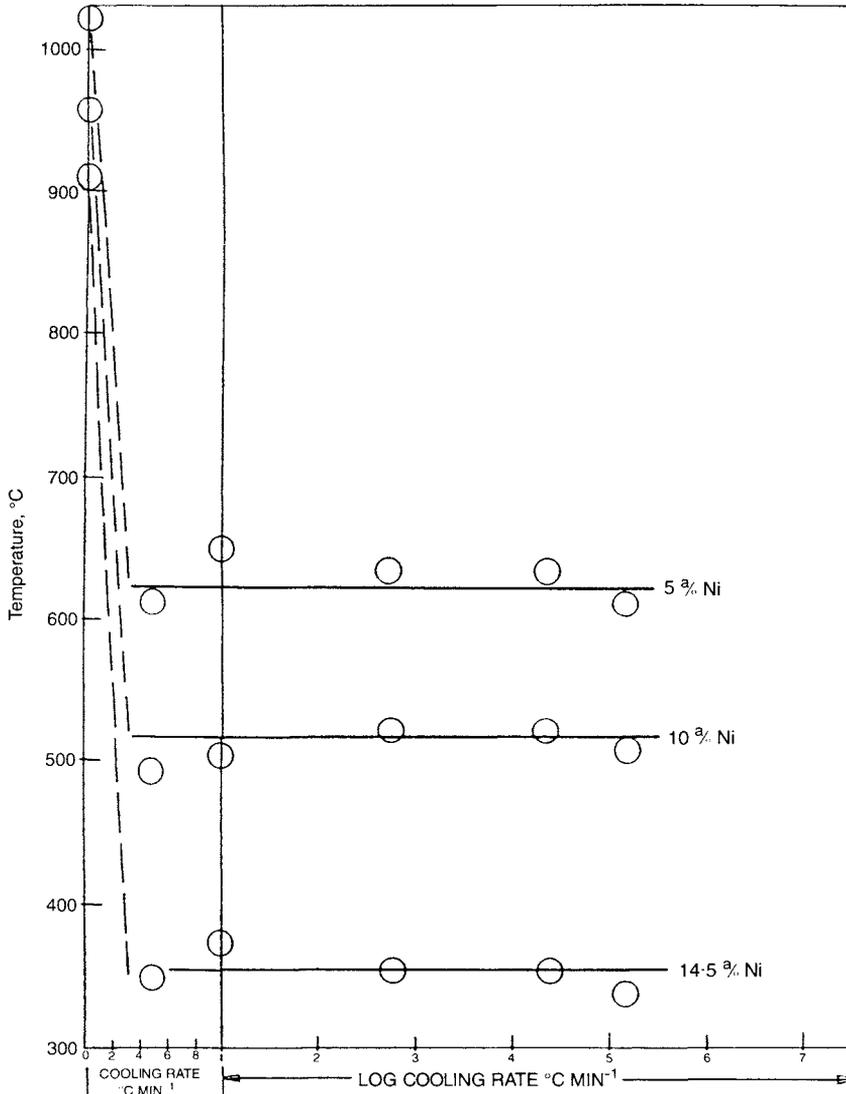


FIG. 15.4. Variation of arrest temperature with cooling rate for iron nickel alloys (after Gilbert and Owen, 1962).

reaction product equi-axed ferrite; it has also been termed grain boundary ferrite. According to Wilson, it nucleates on grain corners which allows the volume change to be accommodated without strain energy. Wilson calls his second product massive ferrite and he believe it nucleates on grain surfaces with a coherent interface with one grain and an incoherent interface which grows into the other grain (the C.S. Smith mechanism). This product cannot form without appreciable strain energy and the additional supercooling is

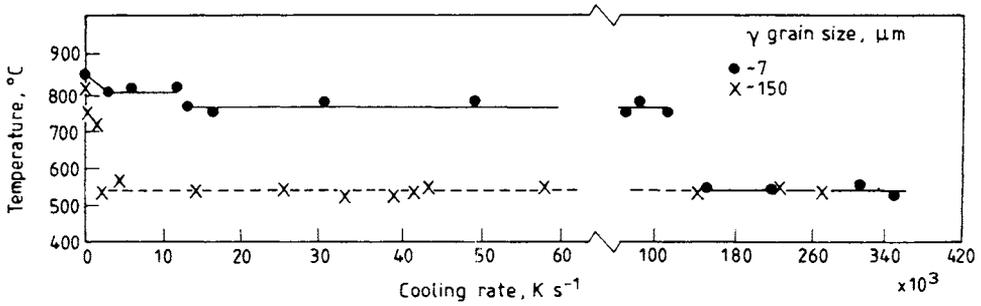


FIG. 15.5. Variation of transformation temperature with cooling rate in zone-refined Fe (after Mirzaev *et al.*, 1987).

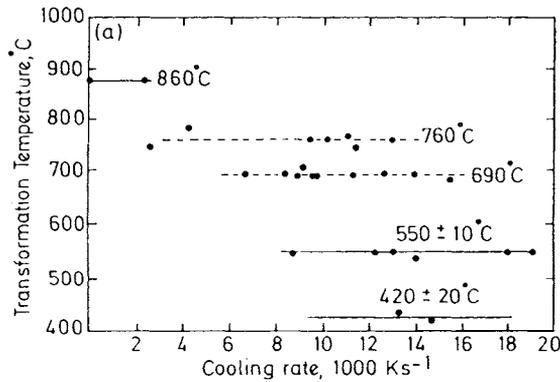


FIG. 15.6. Variation of arrest temperature with cooling rate for iron alloys (after Wilson, 1970, 1984).

necessary to provide the driving force; transformation is now possible only below a temperature  $T'_0$  at which  $Wg^{\alpha\beta} + \Delta g_{cl} = 0$ . Different interpretations have been placed on the structures produced at various cooling rates; for example, Wilson's second massive structure has been described as bainitic and as Widmanstätten ferrite reacted to completion. The Gilbert and Owen results are presumably for the formation of equiaxed ferrite on Wilson's interpretation, but the existence of two types of massive reaction must remain rather controversial; it is consistent with some but not all of the available results.

In the iron alloys, the massive transformation is suppressed not only by very high cooling rates but also by appreciable amounts of substitutional solutes or small amounts of carbon. As shown in Fig. 15.5, the alloying element lowers the temperature of the transformation, and the massive transformation, being dependent on thermally activated growth, can only take place at a temperature where adequate thermal energy is available. This does not apply to martensite unless thermal nucleation is required. Thus at relatively high alloy content, only martensitic reactions occur during rapid cooling.

According to Mirzaev *et al.* (1987) plots of arrest temperature versus cooling rate for other metals (e.g. titanium, zirconium and cobalt) also show several plateaux. Results of this kind for cobalt would be very difficult to understand as the transformation temperature is so low that the transformation is martensitic even at slow rates of cooling.

The formation of a nucleus during continuous cooling was considered by Bhattacharyya *et al.* (1973b). Consider the formation of a nucleus on a  $\beta$ - $\beta$  grain boundary so that a model similar to that of Fig. 10.11 applies. It follows from the shape factors for this model given on p. 456, and from eqn. (46.13), that the number of atoms in a critical nucleus is given by

$$n_c = (16\pi/3)[\sigma^{\alpha\beta}T_o/\Delta h^{\alpha\beta}\Delta T^-]^3 v^2(2 - 3\cos\theta + \cos^3\theta) \quad (71.1)$$

where  $v$  is the atomic volume, for simplicity assumed to be the same in both phases, an expression of type (49.13) has been used to relate the driving force  $\Delta g^{\alpha\beta}$  to the latent heat per atom,  $\Delta h^{\alpha\beta}$ , and the supercooling,  $\Delta T^- = T_o - T_a$ , and  $T_a$  is the temperature at which the nucleus is formed.

An estimate of the time needed to form an embryo of given size may be made by the methods described in Section 50. The simplest such estimate (Hillig, 1962) neglects the shrinkage of embryos (which is important near  $n_c$ ) and treats the embryos as nearly perfect absorbers so that the rate of growth of an embryo containing  $n$  atoms and having surface area  $O_n$  is given by

$$dn/dt = \beta O_n q_o = \beta q_o \eta' v^{2/3} n^{2/3} \quad (71.2)$$

where  $\eta'$  is an appropriate shape factor,  $\beta$  is an accommodation (sticking) coefficient and  $q_o$ , as in Chapter 10, is the jump frequency per unit area of interface from the matrix into the embryo. From the shape factors  $\eta^{\beta}$  and  $\eta^{\alpha\beta}$  it follows that

$$\eta' = 4\pi(1 - \cos\theta)(3/2\pi)^{2/3}(2 - 3\cos\theta + \cos^3\theta)^{-2/3} \quad (71.3)$$

In a continuous cooling experiment, the temperature time curve may be regarded as a series of infinitesimal isothermal treatments for times  $dt$  at temperatures  $T$ , where  $dT/(dT/dt) = dt = dn/(dn/dt)$ . Substituting into eqn. (71.2) gives

$$dn/n^{2/3} = \beta q_o \eta' v^{2/3} dT/(dT/dt) \quad (71.4)$$

As a first approximation, consider that the jump rate does not change significantly in the temperature range from  $T_o$  to  $T_a$ ; this will be valid only for very small values of  $\Delta T^-$ . Integrating eqn. (71.4) between limits  $n = 1$ ,  $T = T_o$  and  $n = n_c$ ,  $T = T_a$  gives

$$n_c^{1/3} \cong n_c^{1/3} - 1 = \beta q_o \eta' v^{2/3} (T_o - T_a)/3(dT/dt) \quad (71.5)$$

where it is assumed, as is usually the case, that  $n_c \gg 1$ . Eliminating  $n_c$  between eqns. (71.1) and (71.5) gives a relation between supercooling and cooling rate of the form

$$dT/dt = K(\Delta T^-)^2 \quad (71.6)$$

A better approximation is obtained if the temperature dependence

$$q_o v^{2/3} = v(\exp \Delta g^*/kT) \quad (71.7)$$

is included, where  $\nu$  is the atomic frequency of vibration and  $\Delta_a g^*$  is the activation energy for jumps across the interface. The  $\alpha$ - $\beta$  interface is assumed to be incoherent so that, in accordance with the discussion on pp. 482–483, the notation  $\Delta_a g^*$  has been used in place of  $\Delta_a g^{\neq}$  shown in Fig. 10.5. Substituting eqn. (71.7) into eqn. (71.4) and integrating between the same two limits gives, by using the result  $\int(e^x/x)dx = e^{-x}(x^{-2} - x^{-1})$  valid for  $x \gg 1$ ,

$$n_c^{1/3} = -(k\beta\nu\eta'/3\Delta_a g^* T)[(T_o \exp -\Delta_a g^*/T_o) - T_a \exp -(\Delta_a g^*/kT_a)] \quad (71.8)$$

Finally, eliminating  $n_c$  between eqns. (71.1) and (71.8) gives

$$dT/dt = K' F(T_o, T_a) \quad (71.9)$$

where  $K' = -k\beta\nu\Delta h^{\alpha\beta}(1 - \cos \theta)/\Delta_a g^* \sigma^{\alpha\beta} T^{\alpha\beta} \nu^{2/3}(2 - 3 \cos \theta + \cos^3 \theta)$  and  $F(T_o, T_a)$  is the function in square brackets in eqn. (71.8).

Bhattacharyya *et al.* applied this theory to the results of Hawbolt and Massalski on a Ag–24 at.% Al alloy. The maximum undercooling observed at cooling rates of up to  $10\,000^\circ\text{C s}^{-1}$  was less than  $100^\circ\text{C}$  so that the approximate theory was considered adequate. A plot of  $\log \Delta T^-$  versus  $\log(dT/dt)$  gave a straight line of slope 1/2 as predicted by the theory, except for the two points at greatest cooling rates. A similar plot could not be used for a Cu–20 at.% Ga alloy because  $T_o$  was not known accurately, but a plot of  $T_o$  versus the square root of the cooling rate gave a linear plot as predicted.

Although this treatment refers specifically to transformation on cooling, it may be equally applicable to changes on heating. In one of the few investigations of the variation of arrest temperature with heating rate, Haworth and Parr (1965) found that superheating of the  $\alpha + \gamma$  change in iron increases to  $50^\circ\text{C}$  as the heating rate is increased to  $10\,000^\circ\text{C s}^{-1}$ . Their data plotted as a log–log relation of superheat versus heating rate could reasonably be fitted to a straight line of slope 1/2 as expected from eqn. (71.6). For the reverse transformation on cooling, values of  $\Delta T^-$  of up to  $180^\circ\text{C}$  are found for cooling rates of up to  $10\,000^\circ\text{C s}^{-1}$  (Duwez, 1951) so that the approximate theory could not be applied. From the data at low values of  $\Delta T^-$ , however, a log–log plot was used to derive an approximate value of  $k$  and this was used to estimate  $K'$  and the relation between  $T_a$  and  $dT/dt$  represented by eqn. (71.9). Bhattacharyya *et al.* found that reasonable agreement with experiment could be obtained with values of  $\Delta_a h^* \cong \Delta_a g^*$  varying from 0.9 to 2.7 eV/atom, so it was not possible to estimate the activation energy for growth with any accuracy.

In this work, the absolute value of the nucleation rate was not considered but later work by Aaronson and his associates has raised grave doubts concerning the reality of the double lens nucleation model.

The early stages of the massive transformations in copper–aluminium, titanium–silver and titanium–gold alloys have been studied by Plichta and Aaronson (1980) and by Plichta *et al.* (1980) respectively. In the case of the Al–Cu particles, all the product crystals (or rather 47 out of 48 of them) formed on grain boundaries, edges or corners with a close approximation to a Burgers orientation relation with one of the adjacent grains. It is assumed that this permits a low energy interface between the two structures. Most of the product in the titanium alloys had low index, rational planes of the parent parallel to the

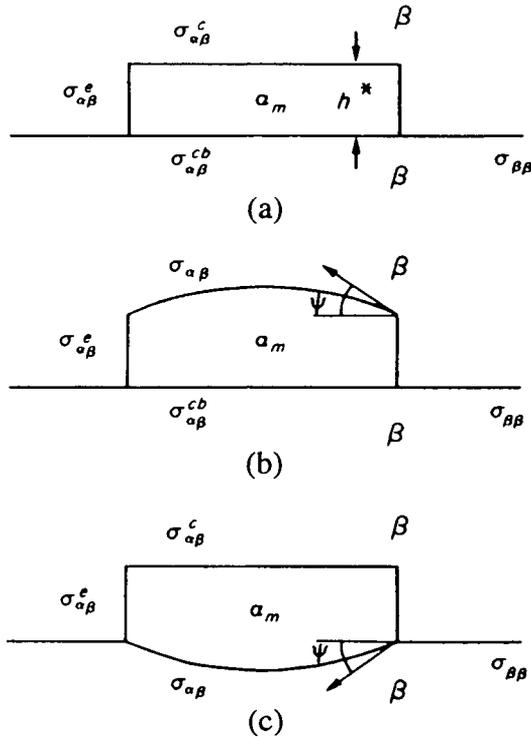


FIG. 15.7. Model of grain boundary nucleation (after Lange and Aaronson, 1979).

habit plane facets and it is similarly assumed that these represent low energy configurations; some calculations with O-lattice theory are offered in support. Both sets of results are interpreted in terms of classical heterogeneous nucleation theory, but with a nucleus shaped like a pillbox (see Fig. 15.7) rather than the lenticular shape discussed in Chapter 10; this enables the configuration to take maximum advantage of the low energy interface. Calculations by the authors indicate a factor of many thousands between the nucleation rate with incoherent interfaces and that with coherent interfaces. Coherent nuclei are able to predict the observed nucleation rate; incoherent nuclei are not. However, a suggestion that the lowest energy nucleus might have a coherent interface with both grains scarcely seems feasible, as the two grains of the parent phase presumably have a random mutual orientation.

The existence of short segments of planar interface was mentioned above. This effect was studied by Hull and Garwood (1956) in a copper-zinc alloy containing 38.7% zinc. They found that the massive transformation could be partly suppressed at fast cooling rates, only a small amount of the massive product being formed near the grain boundaries of the high temperature phase. Single-surface analysis showed that about 40% of the straight segments of the interphase interfaces were consistent with a habit plane of type

{155}–{166} as found in the martensitic transformation in the same alloy at subzero temperatures. The remaining traces could be accounted for as 30% {110}, 20% {112} and 10% {100}, although it should also be noted that about half of the planes ascribed to the {155}–{166} habit could equally well have been described as {112}. The significance of this result and of the observed planar facets is not clear. The incoherent growth mechanism would not be expected to lead to martensitic-type interfaces, but if interfaces of a particular orientation develop to satisfy some condition for coherency, they may persist simply because they are unable to grow rapidly. This explanation of the growth form is similar to that used for crystals growing from the vapour phase; the observed faces are the slow-growing faces. There is the alternative possibility that the transformations may be mixed massive and martensitic in character if the driving force for the two reactions is not very different, so that the grains which begin growing by thermally activated movement across incoherent interfaces may end by growing martensitically at lower temperatures. Certainly at extreme cooling rates there are circumstances in which massive and martensitic products form in the same  $\beta$  grain. However, it does not seem probable that most of the observed planar interfaces can be accounted for in this way. In Massalski's experiments on copper–gallium, copper–zinc, copper–zinc–gallium and copper–gallium–germanium alloys, the massive product was not in contact with untransformed  $\beta$  phase so that habit plane determination could not be made. Long straight boundaries between different product grains were observed, however, and were shown to cross the grain boundaries of the high temperature phase. Massalski suggested that these are growth twin boundaries of the product phase, but no direct confirmation of this has been reported.

Very fast cooling rates from the liquid phase (splat quenching or melt spinning) affect the microstructure obtained. Thus in iron nickel alloys, where the structure at room temperature was martensitic after conventional quenching and after splat quenching, the structure of melt-spun alloys was found to be massive (Hayzelden and Cantor, 1985). The effect is ascribed to a combination of the very small grain size, which suppresses the martensitic nucleation to lower temperatures and thus permits the massive reaction to take place over an extended temperature range, and the peculiar cooling rate associated with melt spinning, which changes from very fast to only moderate below the temperature at which the foil is lifted off the spinning drum.

#### REFERENCES

- BECKER, J. H. (1958) *J. Appl. Phys.* **29**, 1110.  
BHATTACHARYYA, S. K., PEREPEZKO, J. H., and MASSALSKI, T. B. (1973a) *Scr. Metall.* **7**, 485; (1973b) *Ibid.* **7**, 495.  
BIBBY, M. J. and PARR, J. G. (1964) *J. Iron Steel Inst.* **202**, 100.  
BURGERS, W. G. and GROEN, L. J. (1957) *Discuss. Faraday Soc.* **23**, 183.  
CAGLE, F. W. and EYRING, H. (1953) *J. Phys. Chem.* **57**, 942.  
DUNNING, W. J. (1957) *Discuss. Faraday Soc.* **23**, 222.  
DUWEZ, P. (1951) *Trans. Metall. Soc. AIME* **191**, 765.  
EWALD, A. W. and KOHNKE, E. E. (1955) *Phys. Rev.* **97**, 607.  
EWALD, A. W. and TUFTE, O. N. (1958) *J. Appl. Phys.* **29**, 1007.  
GAUNT, P. and CHRISTIAN, J. W. (1959) *Acta Metall.* **7**, 534.

- GILBERT, A. and OWEN, W. S. (1962) *Acta Metall.* **10**, 45.
- GRENINGER, A. B. (1939) *Trans. Metall. Soc. AIME* **133**, 204; (1942) *Trans. Am. Soc. Met.* **30**, 1.
- GROEN, L. J. and BURGERS, W. G. (1954) *K. Ned. Akad. Wet. Proc.* **B57**, 79.
- HALL, E. O. (1956) *The Mechanism of Phase Transformations in Metals*, p. 129, The Institute of Metals, London.
- HAWBOLT, E. B. and MASSALSKI, T. B. (1971) *Metall. Trans.* **2**, 1771.
- HAWORTH, W. C. and PARR, J. G. (1965) *Trans. Mater. Soc. Met.* **38**, 476.
- HAYZELDEN, C. and CANTOR, B. (1985) *Int. J. Rapid Solidif.* **1**, 237.
- HILLIG, W. (1962) *Nucleation and Solidification in Glasses and Metals*, p. 77, American Ceramic Society, Columbus, OH.
- HUBBARD, J. N., MASON, K., and CHRISTIAN, J. W. (1959) *Unpublished work*.
- HULL, D. and GARWOOD, R. D. (1956) *The Mechanism of Phase Transformations in Metals*, p. 219, The Institute of Metals, London.
- LANGE, W. F. III and AARONSON, H. I. (1979) *Metall. Trans.* **10A**, 1951.
- MASSALSKI, T. B. (1958) *Acta Metall.* **6**, 243.
- MIRZAEV, D. A., SCHASTLIVTSEV, V. M., and KURZUMOV, S. Y. (1987) *Fiz. Met. Metalloved.* **63**, 764.
- MOROZOV, O. P., MIRZAEV, D. A., and SHTEINBERG, M. M. (1971) *Phys. Met. Metallogr.* **32**, 127; (1972) *Fiz. Metal. Metalloved.* **34**, 114.
- NEWKIRK, J. B. (1957) *Trans. Metall. Soc. AIME* **209**, 1214.
- PLICHTA, M. C. and AARONSON, H. I. (1980) *Acta Metall.* **28**, 1061.
- PLICHTA, M. C., PEREPEZKO, M. H., AARONSON, H. I., and LANGE, M. M. (1980) *Acta Metall.* **28**, 1031.
- SMITH, R. W. and RAYNOR, G. V. (1957) *Proc. Phys. Soc. London, Sect. B* **70**, 1135.
- WILSON, E. A. (1970) *Scr. Metall.* **4**, 309; (1984) *Met. Sci.* **18**, 471; (1994) *ISIJ Int.* **34**, 615.

## CHAPTER 16

# *Precipitation from Supersaturated Solid Solution*

### 72. TYPES OF PRECIPITATION: SPINODAL DECOMPOSITION

Few topics in physical metallurgy have attracted more attention than the phenomena accompanying the precipitation of a new solid phase from a supersaturated solid solution. The reaction is the basis of the commercial process of age-hardening, and research work on precipitation has been prominent ever since the connection was first recognized. This work has been greatly stimulated by the technological importance of the age-hardening process, but it has also proved to be of great intrinsic scientific interest. In the processes which give age-hardening, usually at relatively low temperatures, there are many complex changes before the formation of the equilibrium structure, and the details often differ between the various alloys which have been investigated.

The precipitation reaction (supersaturated  $\alpha$ )  $\rightarrow \alpha + \beta$  is usually produced by first annealing the specimen at fairly high temperature where its equilibrium structure is single-phase  $\alpha$  and then quenching to a lower temperature where its equilibrium structure is two-phase. The course of the subsequent isothermal reaction may then be followed by measurement of any suitable mechanical or physical property (e.g. hardness or electrical conductivity), but the correlation between the structure of the alloy and measured properties is often very complex. For this reason, direct structural observations by optical or electron microscopy, X-ray analysis and other techniques are most important as a means of understanding precipitation.

In age-hardening alloys, large increases in hardness often occur before any change is visible in the optical microstructure and, when visible precipitates have been formed, the alloy is often found to have softened again, i.e. to be "overaged". The relation between the microstructure and mechanical and physical properties is a very important topic, but is beyond the scope of the present chapter.

The basic requirement for a precipitation reaction is the existence of a solid solution of limited solubility which increases with increasing temperature. The results of all investigations show that it is useful to distinguish the following modes of precipitation:

- (a) spinodal decomposition;
- (b) continuous precipitation;

- (c) discontinuous precipitation;
- (d) Guinier–Preston zone formation and other low temperature phenomena.

The low temperature modes of precipitation are listed separately mainly because direct observation of the product structures is only possible if imaging techniques of very high resolution are used, and also because these processes are often critically dependent on vacancy supersaturation.

Consider a transformation in a binary alloy in which a supersaturated ( $\alpha'$ ) phase breaks up into equilibrium  $\alpha$  phase and a  $\beta$  precipitate. For definiteness let the concentration of solute atoms in the metastable  $\alpha$  phase ( $c^{\alpha'}$ ) be less than that in the precipitate ( $c^{\beta}$ ) and greater than that in the equilibrium  $\alpha$  phase ( $c^{\alpha}$ ). During continuous precipitation, the reaction is proceeding simultaneously in all parts of the assembly although its rate may show considerable variations in different local regions. Isolated crystals of  $\beta$  are nucleated and, as they grow, solute in the surrounding matrix is drained away until the concentration reaches  $c^{\alpha}$ . The original solid solution crystals retain their identities (orientations and external shapes) throughout the reaction. Discontinuous precipitation, in contrast, implies the division of the assembly into regions which have transformed into either equilibrium or non-equilibrium  $\alpha$  and  $\beta$ , and  $\alpha'$  regions of untransformed, supersaturated  $\alpha'$ . In the terminology used in Section 55, the transformed regions are called cells.

Discontinuous precipitation is a process requiring the nucleation and growth of duplex cells. The distribution of precipitates throughout the assembly is non-uniform, and the concentration of solute in the  $\alpha$  phase changes abruptly in a narrow region near the boundary of a cell. At any one time, transformation is taking place at the cell boundaries and the cells are growing into the untransformed  $\alpha'$ . Whereas the kinetics of continuous precipitation depend upon the growth rate of individual  $\beta$  regions, those of discontinuous precipitation depend upon the growth rate of the cells as a whole. The distinction is important: in continuous precipitation, the growth is parabolic and the theory of Section 54 applies but, in discontinuous precipitation, a steady state is attained giving a linear growth rate in accordance with the theory of Section 55.

A further distinction is often made between continuous precipitation which is “general” and that which is “localized”. The distribution of precipitates is uniform in general precipitation, but is confined to particular regions, e.g. grain boundaries of the parent phase, in localized precipitation. Discontinuous precipitation is almost always localized, usually spreading into the grains from the grain boundaries. It may also be produced in single crystals, in which case it must originate either on the free surface or on internal lattice defects.

It is found experimentally that neither the  $\alpha$  nor the  $\beta$  crystals have a systematic orientation relation with the  $\alpha'$  crystals into which they are growing. This results in a sense in discontinuous precipitation involving the fragmentation and recrystallization of the original  $\alpha'$  phase and, for this reason, it has been called the “recrystallization reaction”. The name is misleading as it tends to obscure the fact that the driving force is the difference in chemical free energy between the initial and final conditions. Turnbull (1955) has also

used the term “cellular precipitation” but, although the description above refers to cells of discontinuous precipitate, the generally accepted term discontinuous will be used as the overall description of this form of precipitation.

In some alloys, of which tin and lead are the best known examples, discontinuous precipitation seems to be the only mode of reaction. In other cases there is an initial stage of continuous general or localized precipitation and, under favourable conditions, practically the whole transformation may be effected this way. The reasons for the differences are probably connected with the ease or difficulty of nucleating the two-phase lamellar structure of a discontinuous cell.

Once the cells have formed, they should grow more rapidly than isolated  $\beta$  crystals, as the incoherent boundary provides a good short circuit for the necessary diffusion and the diffusion distance does not increase with time. Discontinuous precipitation is thus more likely to be found under conditions favouring easier nucleation, e.g. low interfacial free energy or high supersaturations to give high driving forces. With high interfacial free energy between the phases, continuous rather than discontinuous precipitation is the probable mode of decomposition. In so far as data are available, they support these predictions.

Evidence of anomalous behaviour in low temperature ageing was first found from the variation of some mechanical and physical properties with ageing time. In some alloy systems, two or more hardening peaks were found and the electrical conductivity first increased and then decreased before any precipitation could be detected by conventional microscopy or X-ray techniques. The details of these first changes differ in different alloys, but it is now established that in many alloys precipitation begins with the formation of clusters rich in solute atoms within the parent lattice.

These clusters retain the structure of the parent phase and are completely coherent with it, although this may produce appreciable elastic strains because the local composition will have a different value within a cluster. The clusters may be detected by X-ray methods or by electron microscopy at an early stage in the process and by atom probe (field-ion) microscopy at a still earlier stage. When they form by localized fluctuations, the clusters are generally called Guinier-Preston (GP) zones, after the two workers who originally discovered them. Clusters also constitute the first stage of spinodal decomposition, although the term GP zones is then not usually used. Indirect evidence from many sources shows that clusters form very rapidly during and immediately after the quench from a high temperature, long before GP zones can be detected.

These effects are intimately connected with excess vacancy concentrations produced by the heat treatments, as discussed later in this chapter.

In the years following the first identification of GP zones, there was much discussion as to whether the zones should be considered as a separate precipitate phase of the assembly or as a pre-precipitation phenomenon in the metastable solid solution. The distinction is largely semantic, especially as zones which are plate-shaped may have thicknesses of only one or two atom diameters, but it is now usual to regard such a zone as a fully coherent precipitate phase. This has the advantage that the formation and dissolution of zones may be interpreted in terms of a metastable equilibrium

diagram in which is plotted the limit of solubility (the GP solvus line) for the coherent phase.

GP zones have different shapes in different alloys, sometimes forming as plates on preferred matrix planes and in other cases as spherical aggregates. Their dimensions are small, typically about 60 atom spacings in diameter, and plate-shaped zones are only one or two atoms thick. The difficulties connected with the unambiguous interpretations of diffuse X-ray patterns led to much controversy in the early work, but it is now accepted that zones form at low temperatures in some but not all precipitation reactions.

The formation of zones is often the first stage in a complex series of changes and may be followed by one or more intermediate transition phases. These are solute-rich phases of definite composition and structure which are never present in the assembly at equilibrium. A metastable structure is usually coherent with the parent phase over most of the interface and is frequently precipitated as plates of definite habit plane.

In many cases, there is insufficient information to decide whether the transitional phase forms directly from the GP zones or is separately nucleated, but evidence for the direct formation of a second precipitate from the zones has been obtained in some cases. In any event, the only reason for the appearance of a non-equilibrium precipitate must be that it can form more rapidly than the equilibrium phase, even though the latter gives a greater reduction in free energy. Of course, if the reaction goes to equilibrium, both the zones and the transitional structure must ultimately disappear and be replaced by the equilibrium phase.

Classical nucleation theory as applied to precipitation reactions is based on the hypothesis that the parent solid solution is metastable to all fluctuations. As already noted, however, if the free energy versus composition curves have the form of Fig. 16.1 there is a limit of metastability inside which the only barrier to the formation of coherent solute-rich and solute-poor regions is the activation energy for atomic diffusion. This unstable region is bounded by the coherent spinodal [eqn. (22.45)] and when an initially homogeneous alloy is cooled rapidly to a temperature inside this region, the solid solution decomposes in a manner quite different from that characteristic of a nucleated transformation. In the remainder of this section, the theory of spinodal decomposition will be examined in more detail.

Becker (1937) and Dehlinger (1937) pointed out that the sign of the diffusion coefficient should be reversed inside the spinodal but the importance of this uphill diffusion in the decomposition of the solid solution was not fully recognized until much later. The first experimental observations of the phenomenon were made by Daniel and Lipson (1943, 1944) who showed that certain anomalous effects ("side-bands") observed in the X-ray diffraction pattern of a Cu-Ni-Fe alloy indicated the existence of a structure with periodic modulations of composition in the  $\langle 100 \rangle$  directions, the modulations having a wavelength of about 10 nm. Borelius (1945, 1951) attempted to develop fluctuation theories of precipitation in which emphasis was laid on the difference between decomposition inside and outside the spinodal. However, these theories were unsatisfactory because they were formulated in terms of localized fluctuations, and the formation of such fluctuations should be difficult because of the accompanying surface energy. A successful theory only

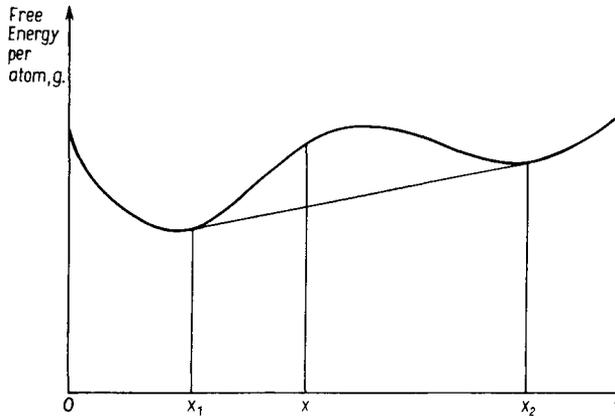


FIG. 16.1. Free energy versus composition curve for a binary system in which the component metals have the same crystal structure and nearly equal atomic sizes.

became possible when it was realized that there are additional energy terms arising from the spatial variation of composition in an inhomogeneous solid solution (Cahn and Hilliard, 1958) and that these terms require a modified form of the diffusion equation which can no longer be adequately described by Fick's law (Cahn, 1961, 1962).

A prediction that spontaneous decomposition inside the spinodal will lead to a periodic composition variation of definite wavelength, as observed by Daniel and Lipson, was first made by Hillert (1961). He used a regular solution model (see Section 23) to discuss one-dimensional diffusion on a discrete lattice, but he modified the flux equation by introducing an energy term arising from the difference in composition of adjacent atomic planes: in effect this is the discrete equivalent of the gradient energy introduced on p. 185. Hillert did not include a coherency energy and his flux equation was non-linear so that the solution had to be obtained numerically. In the three-dimensional continuum model developed by Cahn (1961), the coherency energy was incorporated and an analytical solution to the resulting eqn. (42.12) was obtained by neglecting the non-linear terms in the expansion. This leads to the general solution of eqns. (42.13)–(42.15) which forms the basis of the present theory of spinodal decomposition.

A physical understanding of the theory of spinodal decomposition is aided by recalling the distinction, first made by Gibbs and discussed on pp. 4–5, between large fluctuations from the mean condition confined to localized regions and very small fluctuations extending over large distances. A composition fluctuation of the second kind will tend to disappear again if the diffusion coefficient has its normal sign even if the solid solution is supersaturated, but it will increase in amplitude if uphill diffusion can occur. In this latter case, segregates and clusters will begin spontaneously from any infinitesimal variation in a solid solution of initially uniform composition, and this segregation will eventually result in the formation of the two-phase equilibrium structure.

When the diffusion coefficient is positive, however, the only way in which the two-phase structure can be achieved is through the formation of a large localized fluctuation,

sufficiently different in configuration and/or composition for the net free energy to be lowered. There is then an interface between the localized fluctuation and the solid solution from which it was formed, and the new region can grow further by diffusion of solute in the normal direction, down the composition gradient and towards this interface. A full description of this type of precipitation process must account for both the occurrence of the large fluctuation (“nucleation”) and its subsequent growth by diffusion. The theory of the spinodal process is, in principle, much simpler as only diffusion is involved.

Consider a solid solution of almost uniform composition  $c_0$  but in which there exists an initial infinitesimal fluctuation of arbitrary form which may be represented by

$$c(\mathbf{r}, 0) - c_0 = \iiint A(\boldsymbol{\beta}, 0) \exp(-i\boldsymbol{\beta} \cdot \mathbf{r}) d\boldsymbol{\beta} \tag{72.1}$$

where the wave-vector  $\boldsymbol{\beta}$  has the meaning defined on p. 399 and the coefficients  $A(\boldsymbol{\beta}, 0)$  give the amplitudes of the Fourier components of the fluctuation. At some later time,  $t$ , the composition variation will then be represented by eqn. (42.13) and in particular those components of the fluctuation for which the amplification factor  $R(\boldsymbol{\beta})$  is positive will have increased in amplitude.

Figure 16.2 shows the variation of  $R(\boldsymbol{\beta})$  with  $\beta$ : the amplification factor is negative for values of  $\beta > \beta_c$  [eqn. (42.17)], passes through a maximum and then decreases again towards zero as  $\beta \rightarrow 0$ . By differentiating eqn. (42.15) and equating to zero, it is seen that the maximum value occurs at a wave number  $\beta_m = \beta_c/2^{1/2}$  and is given by

$$R(\beta_m) = 2M_{\text{chem}} K v^2 \beta_m^4 = \frac{1}{2} M_{\text{chem}} K v^2 \beta_c^4 \tag{72.2}$$

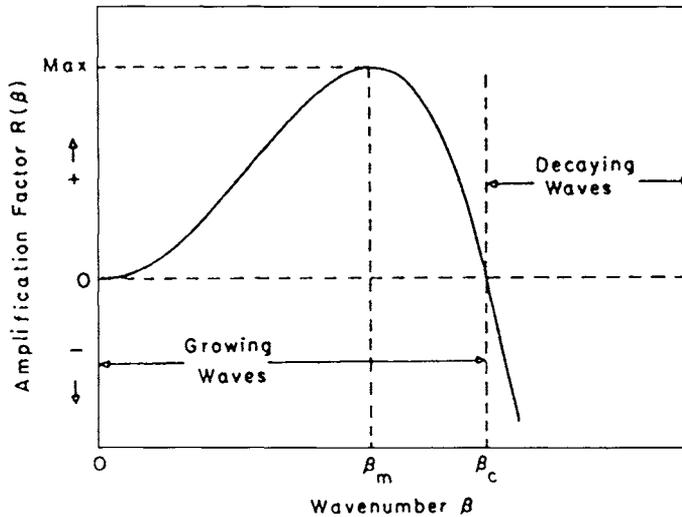


FIG. 16.2. Schematic diagram showing the dependence of the amplification factor  $R(\boldsymbol{\beta})$  on  $\beta$  (after Rundman and Hilliard, 1967).

The negative values of  $R(\boldsymbol{\beta})$  arise because, at small wavelengths  $\lambda < 2\pi/\beta_c$ , the composition gradients are relatively large and the effect of the gradient energy outweighs that of the negative value of  $(\partial^2 g/\partial c^2) + 2v^3 \varepsilon^2 Y'$ . The decrease in  $R(\boldsymbol{\beta})$  at small wave numbers (large wavelengths) is a kinetic effect; segregation is energetically favourable for such components but proceeds slowly because of the relatively large diffusion distances. In fact, for sufficiently small values of  $\beta$ ,  $R(\boldsymbol{\beta})$  varies as  $\beta^2$ , i.e. as  $\lambda^{-2}$ , which is the result obtained from the classical form of the diffusion equation without consideration of a gradient energy.

The coherency strain energy introduced on p.185 has magnitude  $v\varepsilon^2 Y'(x-x_0)^2 = v^3 \varepsilon^2 Y'(c-c_0)^2$  per atom, where  $Y'$  was defined for an elastically isotropic medium. In such a medium, fluctuations are equally probable in all directions and the wavelength of the fastest growing Fourier component does not vary with direction. This applies, for example, to spinodal decomposition in glasses, but all crystalline materials are elastically anisotropic and the effects of the anisotropy have now to be considered.

For a cubic crystal, the stress-free strains correspond to a pure dilatation and the coherency energy was calculated by Cahn (1962) by assuming that, for a plane wave in a given direction, the lattice parameter is held constant at the value corresponding to the mean composition  $c_0$  in all planes normal to the direction of the variation in composition, but that the stress is relaxed to zero in the direction of the fluctuation. This leads to an expression for the coherency energy of the same form as that given above for the isotropic case but with  $Y'$  replaced by a parameter  $Y'(\boldsymbol{\beta})$  which varies with the direction of  $\boldsymbol{\beta}$ . The minimum value of  $Y'$ , and hence of the coherency energy, occurs for composition fluctuations along either  $\langle 100 \rangle$  or  $\langle 111 \rangle$  directions of the lattice, where

$$Y'(100) = c_{11} + c_{12} - 2c_{12}^2/c_{11} \quad (72.3)$$

and

$$Y'(111) = 6c_{44}(c_{11} + 2c_{12})/(c_{11} + 2c_{12} + 4c_{44}) \quad (72.4)$$

The direction of minimum energy depends on whether the anisotropy factor  $2c_{44}/(c_{11} - c_{12})$  is greater or smaller than unity. For most metals, this factor is greater than unity (see Table IV, p. 76) and minimum energy fluctuations are in  $\langle 100 \rangle$  directions but, for materials with an anisotropy factor of less than unity, the preferred directions are  $\langle 111 \rangle$ .

The anisotropic coherency energy has the effect of introducing directional behaviour into diffusion phenomena even in cubic systems and, in addition to spinodal decomposition, this can be important in, for example, the coarsening of a distribution of small coherent precipitate particles. The total free energy of a system with composition fluctuations in a single direction (plane waves) may be expressed as a single volume integral, as in eqn. (42.8), but the general theory of the strain energy for three-dimensional fluctuations is more complex as the contribution of each Fourier component depends on its direction as well as its magnitude. Fortunately, the effects of elastic anisotropy are easily expressed for spinodal decomposition. In the linear approximation, eqn. (42.15) remains valid if  $Y'$  is replaced by  $Y'(\boldsymbol{\beta})$ . It is more difficult to modify the diffusion equation to take full account of the anisotropy, and this problem is now briefly discussed.

The right-hand side of eqn. (42.13) is the three-dimensional Fourier transform of the amplitude spectra and it follows that

$$A(\boldsymbol{\beta}, t) = (2\pi)^{-3} \iiint (c(\mathbf{r}, t) - c_0) \exp(-i\boldsymbol{\beta} \cdot \mathbf{r}) d\mathbf{r} \quad (72.5)$$

The strain energy of a coherent cubic system with composition varying in three dimensions is written by Cahn (1962, 1965) as

$$(2\pi)^3 v \varepsilon^2 \iiint Y'(\boldsymbol{\beta}) A(\boldsymbol{\beta}) A^*(\boldsymbol{\beta}) d\boldsymbol{\beta} \quad (72.6)$$

where  $A^*(\boldsymbol{\beta})$  is the complex conjugate of  $A(\boldsymbol{\beta})$ . By considering the free energy change resulting from a variation  $\delta c$  in the composition at  $r$ , Cahn obtained an equation for the diffusion flux which is similar to eqn. (42.11) but has an integral term derived from eqn. (72.6) in place of  $2\varepsilon^2 Y' v^2$ .

Similarly, the revised form of the differential equation for diffusion, obtained by taking the divergence of the flux, is identical with the isotropic form (42.12) except that the term  $2M_{\text{chem}} \varepsilon^2 Y' v^2 \nabla^2 c$  is replaced by

$$2M_{\text{chem}} \varepsilon^2 v^2 \iiint -\boldsymbol{\beta}^2 A(\boldsymbol{\beta}) Y' \exp(-i\boldsymbol{\beta} \cdot \mathbf{r}) d\boldsymbol{\beta} \quad (72.7)$$

Expression (72.7) involves a convolution integral of the transforms of the individual factors, and if this can be evaluated the diffusion equation is expressed entirely in terms of the composition and its spatial derivatives. However, a solution of the linearized diffusion equation in terms of the components  $A(\boldsymbol{\beta})$  may be obtained without evaluating the integral. As already stated, this solution has the form of eqn. (42.15) with  $Y'$  replaced by  $Y'(\boldsymbol{\beta})$  so that, according to this continuum model, orientation effects in the early stages of spinodal decomposition arise essentially only because the amplification factor  $R(\boldsymbol{\beta})$  contains this orientation-dependent modulus.

Cahn's calculation of the coherency strain energy is valid only for an elastic continuum, and some differences may be expected in a crystal lattice if the wavelength of the composition variation is not appreciably larger than the lattice spacing. Cook and De Fontaine (1969) considered this problem of "atomic elasticity" and obtained two new results, namely that the elastic energy depends on the magnitude as well as the direction of  $\boldsymbol{\beta}$ , and that a crystal which is elastically isotropic on a macroscopic scale may nevertheless be anisotropic on an atomic scale.

The wavelength dependence may be regarded as an elastic contribution to the gradient energy and thus incorporated into the gradient energy coefficient  $K$  of the continuum model; however, this energy is now also orientation-dependent so that  $K$  strictly becomes  $K(\boldsymbol{\beta})$ . In the following description this complication is ignored.

Experimental tests of the revised diffusion equation have been made by the technique of small angle X-ray scattering which, as described below, enables  $A(\boldsymbol{\beta})$  and  $R(\boldsymbol{\beta})$  to be derived directly. In order to describe the course of spinodal decomposition, however, the composition  $c(\mathbf{r}, t)$  must be derived from the inverse Fourier transform of  $A(\boldsymbol{\beta}, t)$  [eqn. (42.13)]. In view of the exponential dependence of  $A(\boldsymbol{\beta}, t)$  on  $R(\boldsymbol{\beta})$ , it is plausible to assume

that components with wave-vectors close to those giving the maximum value of  $R(\boldsymbol{\beta})$  will become dominant after a very short time. An approximate solution is thus obtained by maximizing  $R(\boldsymbol{\beta})$  with respect to both the magnitude and direction of  $\boldsymbol{\beta}$  and supposing that only waves with these maximum values need be considered. For a fixed direction, there is a critical wave number  $\beta_c$  and a wave number  $\beta_m = \beta_c/2^{1/2}$  at which  $R(\boldsymbol{\beta})$  for that direction has its maximum value. The orientation dependence of  $\beta_c$  is obtained from eqn. (42.17) with  $Y'$  replaced by  $Y'(\boldsymbol{\beta})$  and, as the maximum values of  $R(\boldsymbol{\beta})$  vary as  $\beta_c^4$  [eqn. (72.2)], it follows that the rate of growth of fluctuations is indeed very orientation-dependent.

When  $Y'$  becomes  $Y'(\boldsymbol{\beta})$ , eqn. (22.45) will define a directionally dependent coherent spinodal, but the solid solution will be metastable to fluctuation in all directions only outside the curve defined by the minimum value  $Y'_m$  of  $Y'(\boldsymbol{\beta})$ , so that this should be regarded as the operative coherent spinodal. The difference  $T^s - T^{cs}$  in the maximum temperatures of the chemical and coherent spinodal curves may be written approximately in terms of an assumed constant rate of change of  $(\partial^2 g/\partial c^2)$  with temperature, i.e. of a temperature-independent value of  $(\partial^2 s/\partial c^2)$ , where  $s$  is the entropy per atom. Note that  $(\partial^2 s/\partial c^2) = (-1/T)(\partial^2 g/\partial c^2)$  is negative for a system showing spinodal decomposition.

This approximation gives

$$T^s - T^{cs} = -2\varepsilon^2 Y'_m v^3 / (\partial^2 s/\partial c^2) \quad (72.8)$$

Using eqn. (42.17), the maximum value of  $\boldsymbol{\beta}$  is defined by

$$\beta_m^2 = \frac{1}{2}\beta_c^2 = -\{v^{-3}(\partial^2 g/\partial c^2) + 2\varepsilon^2 Y'_m\}/4K \quad (72.9)$$

and again expanding  $(\partial^2 g/\partial c^2)$  linearly about its (zero) value at  $T = T^s$

$$\beta_m^2 = \{(T - T^s)v^{-3}(\partial^2 s/\partial c^2) - 2\varepsilon^2 Y'_m\}/4K \quad (72.10)$$

finally, substituting eqn. (72.8) into eqn. (72.10) gives

$$\beta_m^2 = (T - T^{cs})(\partial^2 s/\partial c^2)/4Kv^3 \quad (72.11)$$

Equation (72.11) shows that  $\beta_m$  increases with increasing supercooling below the coherent spinodal. It then follows from eqn. (72.2) that  $R(\beta_m)$  increases rapidly from zero as the supercooling is increased, but because of the temperature dependence of  $M_{\text{chem}}$  the value of  $R(\beta_m)$  will reach a maximum with decreasing temperature, and then will decrease again. Thus, as pointed out by Huston *et al.* (1966), the isothermal transformation curves for spinodal decomposition will have the familiar C-shape (see p. 547).

An isothermal transformation diagram given by these authors for a "hypothetical but not unrealistic" set of parameters shows the "nose" of the diagram at a temperature  $T/T^{cs} = 0.93$ , with  $\lambda_m = 10$  nm at this temperature. Except for temperatures close to  $T^{cs}$ , the decrease of  $\lambda_m$  with decreasing temperature [eqn. (72.11)] is not rapid, i.e. in the example just cited,  $\lambda_m = 5$  nm at  $T/T^{cs} = 0.75$ .

The preferential development of particular Fourier components of the composition variation has implications for the morphology of a system which has undergone spinodal decomposition. If the material is elastically isotropic, or if  $Y'$  is sufficiently small for coherency strain energy effects to be insignificant, the dominant fluctuations will have a

preferred wavelength but will be in random directions. Cahn (1965) has studied the distribution of solute in this case by computer simulation and there are several experimental studies of spinodally decomposed glasses. Strictly, the linearized theory does not predict a two-phase structure, and the fluctuations are necessarily symmetric about the average composition. Nevertheless it may be approximately valid in the case of equiatomic ( $x_B = \frac{1}{2}$ ) alloys to regard the eventual interphase interfaces as the surfaces of composition  $c^{a'}$  and for other values of  $x$  similarly to regard the interphase interfaces as the isoconcentration surfaces which divide the crystal into the correct volume fraction of the equilibrium phases. The structures so defined do not have any obvious periodicity because of the random directions of the components, but connectivity of the two phases is obtained over a wide range of compositions (in Cahn's simulations, for atom fractions  $0.15 < x_B < 0.85$ ). In contrast, a nucleated reaction begins, of course, by the formation of discrete regions of one phase embedded in the other phase, and these regions are unlikely to coalesce and become interconnected unless the final volume fraction of the precipitate phase is large.

Consider the structure likely to develop in most cubic systems in which the elastically soft directions are  $\langle 100 \rangle$ . The superposition of sinusoidal waves of equal amplitude and wavelength along the three  $\langle 100 \rangle$  directions leads to composition maxima and minima situated at the  $A$  and  $B$  sites respectively of a CsCl (B2) type structure of cube edge  $\lambda$ .

The surfaces of constant concentration  $c_0$  define tetrakaidecahedra (truncated octahedra; see Fig. 8.4) around each maximum and minimum, so that for equal volume fractions of the two phases, the final structure expected is approximately a regular array of tetrakaidecahedral particles of the two phases forming together a CsCl structure of lattice parameter  $\lambda$ . When one phase has a volume fraction appreciably smaller than the other, it will take the form of approximately octahedral particles in a simple cubic array of lattice parameter  $\lambda$ ; the corners of the octahedra will be aligned along the  $\langle 100 \rangle$  directions.

For crystals in which the elastically soft directions are  $\langle 111 \rangle$ , the theory is rather more complex because the mutual interaction of composition fluctuations along four different directions must be considered. Any two  $\langle 111 \rangle$  waves will produce maxima along a regular array of rods aligned in the common  $\langle 110 \rangle$  direction, and the other two waves then result in a periodic variation along these rods. Different choices of the first two wave directions will lead to rods along all six  $\langle 110 \rangle$  directions. The final structure (Cahn, 1964; Hilliard, 1970) is thus an interlaced set of  $\langle 110 \rangle$  rods, each with periodic variation in thickness. The rods are continuous unless (coincidentally) the composition fluctuations, in addition to having the same amplitude, are all in phase. For this special situation, the resulting structure is of the NaCl type, with the maxima and minima occupying the atom sites (Cahn, 1962).

Significant features of these morphological predictions are the fine scale of the structure and, in elastically anisotropic crystals, the alignment of the precipitate particles into a regular array. As discussed below, the scale of the structure is not changed much by changing the temperature of the reaction or by transforming during continuous cooling rather than isothermally, but the spinodal decomposition may be succeeded by a coarsening reaction. In the early days of the theory, the observation of a fine, regularly

aligned precipitate was taken as an indication of spinodal decomposition, but it is now known that the coherency strain fields of particles produced by nucleation and growth can so affect the diffusion flux during a subsequent coarsening reaction that alignment of the resulting larger particles is obtained. Thus the observation of an aligned structure is not sufficient evidence that a spinodal reaction has occurred.

Equations (42.13)–(42.15) have the property that those Fourier components that are unstable will continue to increase indefinitely in amplitude. This is obviously physically unrealistic and arises because the free energy versus composition relations have been characterized in the linearized form of eqn. (42.12) by a single value of  $(\partial^2 g/\partial c^2)$ , so that  $g(c)$  is effectively a parabola and decreases indefinitely as the atomic fraction tends to zero or unity. The real free energy versus composition curve has the general shape of Fig. 16.1 and, as the maximum and minimum compositions of a Fourier component approach the spinodal compositions, the growth of that component must slow down and ultimately cease. In order to produce a free energy versus composition curve with two minima, and thus to predict this slowing down by means of an expansion about  $c = c_0$ , it is necessary to include terms containing at least  $(\partial^3 g/\partial c^3)$  and  $(\partial^4 g/\partial c^4)$ . It follows that the curve of spinodal decomposition developed above will remain valid so long as the composition excursions are confined to a region in which  $(\partial^2 g/\partial c^2)$  can be considered as essentially unchanged from its value at  $c = c_0$  but that, outside this region, the non-linear terms of eqn. (42.12) become increasingly important.

Following Cahn (1965), eqn. (42.12) is written in the form

$$\partial c/\partial t = D_1 \nabla^2 c - D_2 \nabla^4 c + D'_1 (\nabla c)^2 + \dots \quad (72.12)$$

where  $D'_1 = \partial D_1/\partial c$ . In the linear form of the equation, the third term is ignored and  $D_1$  and  $D_2$  are given the values appropriate to  $c = c_0$ . As an alternative to this assumption, each coefficient is now expanded about its value at  $c_0$  to obtain

$$\partial c/\partial t = D_1 \nabla^2 c - D_2 \nabla^4 c + D'_1 \nabla[(c - c_0)\nabla c] + D''_1 \nabla[(c - c_0)^2 \nabla c] \quad (72.13)$$

where the quantities  $D_1$ ,  $D_2$ ,  $D'_1$  and  $D''_1$  are all given their values at  $c = c_0$ . The third term in eqn. (72.13) thus includes the value of  $(\partial^3 g/\partial c^3)$  and the fourth term includes  $(\partial^4 g/\partial c^4)$ .

Cahn (1965) showed how to solve eqn. (72.13) by a method of successive approximations, but only his main conclusions will be mentioned. The non-linear terms give rise to harmonic distortions of the composition waves of the linear solution, and these are of two types. Odd harmonics arise from even derivatives of  $D_1$ , and thus primarily from  $(\partial^4 g/\partial c^4)$ . They decrease the growth rate of the wave, flatten the extremes of composition and increase the gradient between the extremes. Thus a sine-wave fluctuation is converted into a profile approaching a square wave which is more characteristic of a two-phase structure. Even harmonics arise from odd derivatives of  $D_1$ , i.e. primarily from  $(\partial^3 g/\partial c^3)$ , and have a different effect. They enhance the composition differences of one sign more than those of the other sign whilst at the same time narrowing the spatial extent of the enhanced (minor phase) differences in composition. They also tend to break the connectivity of the minor phase. These are exactly the characteristics required to produce a two-phase structure in accordance with the lever rule under conditions where the volume

fractions of the two phases are not equal. A non-zero value of  $(\partial^3 g / \partial c^3)$  means that the free energy versus composition curve is not symmetrical about  $c = c_0$  and thus that the volume fractions of the two phases are not the same.

It is a rather remarkable result of Cahn's theory that the predictions are physically correct even for the later stages of decomposition, despite the obvious limitations of the model. However, in order to obtain his approximate solutions, he made the simplifying assumption that fluctuations of a single wavelength develop in the early stages. Whilst this is consistent with the linear theory as developed above, Hilliard (1970) has pointed out that experimental results indicate that the actual spectrum is much less sharply peaked about the wavelength  $\lambda_m$  corresponding to maximum growth rate. A more elaborate non-linear theory has been developed by Langer (1973) but the details are too complex to be included here.

Although it has been emphasized above that the linear theory should apply only to the early stages of the transformation, it is now necessary to mention an opposite difficulty, namely that the linear theory also does not strictly apply at the very beginning of the transformation because of the neglect of thermal fluctuations. Consider, for example, a temperature and composition where the single-phase solid solution is stable. According to eqn. (42.14),  $A(\beta, t)$  decreases rapidly to zero for all  $\beta$  so that, whatever initial composition distribution is assumed, a uniform composition is achieved and thereafter there is no diffusion flux. However, this is only an average or macroscopic view of the equilibrium solid solution; on an atomic scale, there is continual rearrangement, with small clusters of solute atoms, small ordered regions, etc. continually forming and dispersing again (see Sections 27 and 47). This dynamic equilibrium means that the instantaneous flux of solute at any point, as distinct from its time average, is never zero. Cook (1970) pointed out that the motion of an individual solute atom is strictly analogous to the phenomenon of Brownian motion of small solid particles suspended in a liquid.

The Brownian motion of the solute in a metastable solid solution is central to the mechanism of nucleation by a series of successive fluctuations. Its effects inside the coherent spinodal are much less significant, but they cannot be entirely ignored in the very early stages of the reaction. When the initial situation is very close to a uniform distribution of solute, the Fourier amplitudes of the composition modulation will not grow or shrink with time monotonically as predicted by eqn. (42.14), but will have a superimposed oscillatory growth/shrinkage due to random thermal fluctuations. Cook showed that this effect can be ascribed to a thermal change force proportional to  $RT$ . Brownian motion is further discussed below in relation to experimental investigations of spinodal decomposition.

The treatment above has been developed for spinodal decomposition at constant temperature but for many spinodal reactions the combination of high atomic mobility and short diffusion distance at the temperature of the spinodal means that it is impossible to prevent some decomposition, however fast the material is quenched. This may be illustrated by eliminating  $K$  between eqns. (72.2) and (72.9) to give

$$R(\beta_m) = -\frac{1}{2} M_{\text{chem}} \{v^{-1} (\partial^2 g / \partial c^2) + 2\varepsilon^2 Y' v^2\} \beta_m^2 \quad (72.14)$$

As pointed out by Huston *et al.* (1966), eqn. (72.14) may be regarded simply as a version of the basic diffusion law (see p. 389) as the diffusion distance  $\lambda_m = 2\pi/\beta_m = (Dt)^{1/2}$ , where the time of the isothermal reaction is given by  $t \approx 1/R(\beta_m)$ . Typical values of  $D$  for non-viscous liquids and for solids near their melting points are  $10^{-9} \text{ m}^2 \text{ s}^{-1}$  and  $10^{-13} \text{ m}^2 \text{ s}^{-1}$  respectively, and a typical value of  $\lambda_m$  is 10 nm. This gives isothermal reaction times of the order of microseconds or less for liquids undergoing spinodal decomposition and of the order of milliseconds for solids near the melting point.

Thus it is likely to be possible to cool sufficiently rapidly to prevent transformation, and subsequently to study isothermal transformation, only in alloy systems where the maximum temperature of the coherent spinodal is appreciably lower than the melting point.

On p. 726 it was mentioned that the wavelength  $\lambda_m$  is a weak function of the temperature of isothermal reaction, except for very small supercooling from the spinodal temperature  $T^{\text{cs}}$ . It follows from this that structures produced by transformation during continuous cooling will not be very different from those already discussed. However, because the wavelength of the fastest-growing component will decrease somewhat as the reaction proceeds (and the temperature falls), the final distribution of composition will correspond to a broader spread of wavelengths in the vicinity of  $\lambda_m$  than is the case for isothermal decomposition.

Huston *et al.* have given a quantitative treatment of these effects. They assumed that  $\varepsilon$ ,  $Y'$  and  $K$  are constants, or at least have a temperature variation which is insignificant compared with that of  $(\partial^2 g/\partial c^2)$  and  $M_{\text{chem}}$ . During the cooling, a Fourier component of particular wave number  $\beta$  will begin to grow at a temperature  $T_\beta$  satisfying the equation

$$v^{-3}(\partial^2 g/\partial c^2)_{T=T_\beta} + 2\varepsilon^2 Y' + 2K\beta^2 = 0 \quad (72.15)$$

and the amplitude of this component when temperature  $T$  is reached is obtained by combining eqns. (42.14) (42.15) and writing the resultant equation in the temperature-dependent integral form

$$A(\beta, T) = A(\beta, T_\beta) \exp \left[ \int_{T_\beta}^T -M_{\text{chem}} \beta^2 \frac{\{v^{-1}(\partial^2 g/\partial c^2) + 2\varepsilon^2 Y' v^2 + 2Kv^2 \beta^2\}}{(\partial T/\partial t)} dT \right] \quad (72.16)$$

The temperature-dependent quantities in the integral are  $(\partial^2 g/\partial c^2)$ , which is assumed to vary linearly with  $T - T^{\text{cs}}$  (see p. 726), and  $M_{\text{chem}}$ , which has an Arrhenius type of temperature variation, the activation energy being  $\Delta h_0 + \Delta_a h_0$  if the cooling rate  $-dT/dt$  is sufficiently slow to maintain the concentration of vacancies in thermal equilibrium, and  $\Delta_a h_0$  if the cooling rate is sufficiently rapid to prevent any vacancies disappearing during the quench. Huston *et al.* evaluated the integral for the limiting cases in which the cooling rate is sufficiently rapid to prevent any transformation and sufficiently slow to allow complete transformation. Figure 16.3 shows for a hypothetical system their results for the dependence of the wave number  $\beta_m$  receiving maximum amplification on the cooling rate. In the region where almost complete decomposition occurs during cooling,  $\beta_m$  increases (i.e.  $\lambda_m$  decreases) with the one-sixth power of the cooling rate: there is also an increase in

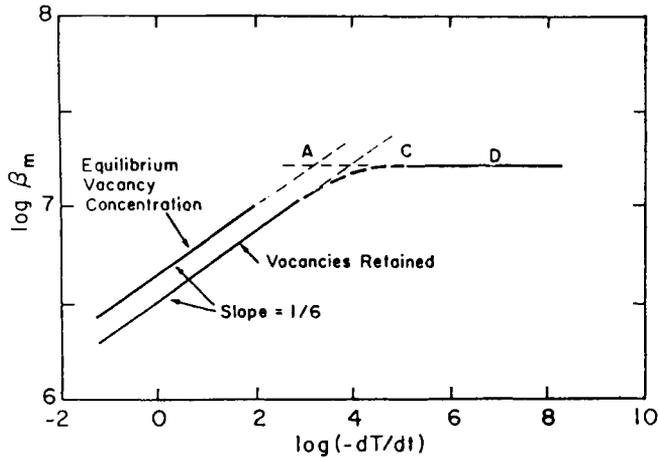


FIG. 16.3. Plot of logarithm of wave number receiving maximum amplification versus the logarithm of the cooling rate (after Huston *et al.*, 1966). The broken portions of the curves represent the change from the slow to the fast cooling rates. For cooling rates to the right of point *C*, decomposition is incomplete.

$\lambda_m$  at fixed cooling rate if it is assumed that the rate is sufficiently fast to maintain the original concentration of vacancies. This is one example of a reaction where non-equilibrium vacancy concentrations obtained through rapid cooling effects the process of phase transformation by providing enhanced diffusivity. Effects of this kind can only be observed experimentally when the diffusion distance is small in comparison with the distance of the average vacancy from a sink.

For sufficiently fast cooling rates, to the right of point *C* in Fig. 16.3, decomposition is not completed during the cooling, and the wavelength of maximum amplification becomes independent of cooling rate. At cooling rates greater than that represented by point *D*, there is essentially no growth of an initial fluctuation during the whole time taken for the quench.

Figure 16.4 shows a continuous cooling transformation diagram, again calculated for a hypothetical system. Linear cooling curves denoted  $Q_D$  and  $Q_C$  in this figure correspond to the points *D* and *C* in Fig. 16.3 and, for all cooling rates less than  $Q_D$ , the intersections of the temperature versus time curves with the 1% and 99% transformation lines effectively define the temperatures and times at which the transformations begin and end. The temperature of the 99% completion curve also defines the wavelength  $\lambda_m$  which can be read from the right-hand vertical scale on the same diagram.

When it is possible to suppress spinodal decomposition by rapid cooling, the kinetics of a subsequent isothermal reaction may be used to test the above theory. Such experiments are primarily tests of the modified diffusion equation and are related to the experiments designed to verify this equation, in non-spinodal situations, as mentioned on p. 400.

The work of Daniel and Lipson first showed that composition modulations in a solid solution lead to side-band phenomena, i.e. to satellite maxima around the Bragg peaks in

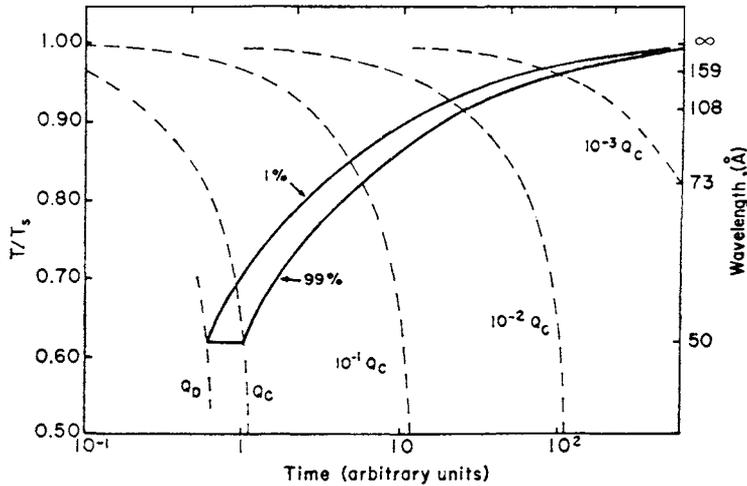


FIG. 16.4. Continuous cooling diagram calculated for a hypothetical system (after Huston *et al.*, 1966).

an X-ray diffraction experiment. Because of the combined effects of the spatial variations of X-ray scattering power and of lattice parameters, the theory of the distribution of intensity is very complex, but the intensity distribution near the origin of reciprocal space, i.e. at small scattering angles from the incident beam, is relatively easy to handle because the contribution from varying lattice parameter is then negligible. This was first realized by Rundman and Hilliard (1967), who demonstrated in a classical experiment the existence of spinodal decomposition in aluminium zinc alloys.

Consider a point  $\mathbf{s}$  in reciprocal space where the vector  $\mathbf{s}$  is normal to the lattice planes of spacing  $d_s = 1/|\mathbf{s}|$  [see eqn. (5.22)]. X-rays of wavelength  $\lambda_s$  diffracted in the direction  $\theta$  may then be represented by the scattered amplitude at point  $\mathbf{s}$  where (satisfying the Bragg equation)  $|\mathbf{s}| = (2 \sin \theta)/\lambda_s$ . This scattered amplitude is given by

$$A(\mathbf{s}, t) = \int_V P(\mathbf{r}, t) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r}) d\mathbf{r} \quad (72.17)$$

where  $P(\mathbf{r}, t)$  is the electron density at point  $\mathbf{r}$  and time  $t$  and is related to the composition fluctuation through

$$P(\mathbf{r}, t) = P_0 + (f_B - f_A)[c(\mathbf{r}, t) - c_0] \quad (72.18)$$

In eqn. (72.18)  $P_0$  is the mean electron density and  $f_A$  and  $f_B$  are the atomic scattering factors which are effectively constant in the small angle region. From eqns. (72.17) and (72.18)

$$A(\mathbf{s}, t) = (f_B - f_A) \int_V [c(\mathbf{r}, t) - c_0] \exp(-2\pi i \mathbf{s} \cdot \mathbf{r}) d\mathbf{r} \quad (72.19)$$

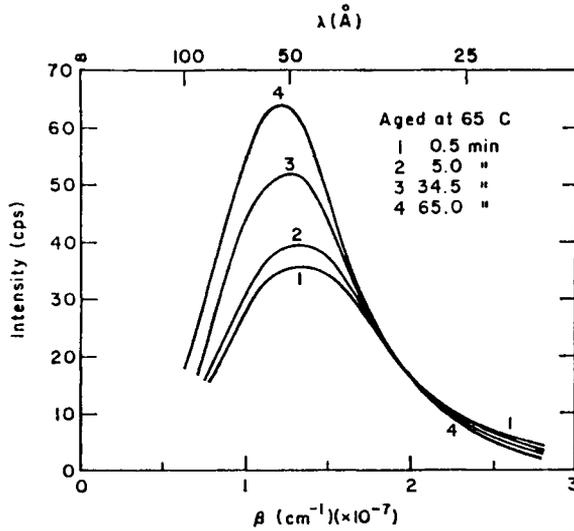


FIG. 16.5. Small angle X-ray spectra for an Al 22 at% Zn alloy quenched from 425 C and annealed at 65 C for the times indicated (after Rundman and Hilliard, 1967).

Equation (72.19) may be compared with eqn. (72.5) and thus leads directly to an equation of the form (42.14). However, measurements can only be made of the scattered X-ray intensity  $I(\mathbf{s}, t) = I(\boldsymbol{\beta}, t)$  given by the product of the scattered amplitude with its complex conjugate, so that this equation has to be written

$$I(\boldsymbol{\beta}, t) = I(\boldsymbol{\beta}, 0) \exp[2R(\boldsymbol{\beta})t]; \quad \beta = 2\pi s = 4\pi \sin \theta / \lambda_v \quad (72.20)$$

Thus at values of  $\beta < \beta_c$ , the X-ray intensity  $I(\boldsymbol{\beta}, t)$  should increase with time in the early stages of decomposition whereas, for values greater than  $\beta_c$ , it should decrease. Figure 16.5 shows the results obtained by Rundman and Hilliard for an Al-Zn alloy with  $X_{Zn} = 0.22$  which was quenched from 425 C and then aged at 65 C.

The curves display the two features expected from eqn. (42.15), namely they have a common cross-over point (corresponding to  $\beta_c$ ) and each curve has a maximum at a common value ( $\beta_m$ ) of  $\beta$ . The values of  $R(\boldsymbol{\beta})$  were obtained from the slopes of the linear plots of  $\ln I(\boldsymbol{\beta}, t)$  versus  $t$ , and a derived plot of  $R(\boldsymbol{\beta})/\beta^2$  versus  $\beta^2$  was then also found to be linear (see Fig. 16.6). It follows from eqns. (42.15) and (42.5) that the slope of this derived plot is equal to  $-2M_{\text{chem}}Kv^2 = -2D_{\text{chem}}Kv/(\partial^2 g/\partial x_B^2)$  and its intercept on the  $\beta = 0$  axis is  $-D_{\text{chem}}\{1 + (2\varepsilon^2 Y'v)/(\partial^2 g/\partial x_B^2)\}$ . The quantity  $R(\boldsymbol{\beta})/\beta^2$  may be regarded as defining an effective diffusion coefficient  $D(\boldsymbol{\beta})$  where

$$D(\boldsymbol{\beta}) = -R(\boldsymbol{\beta})/\beta^2 = D_{\text{chem}} \{1 + 2v\varepsilon^2 Y' + K\beta^2(\partial^2 g/\partial x_B^2)^{-1}\} \quad (72.21)$$

In the particular case of Al-Zn alloys, the coherency energy is small because of the small value of  $\varepsilon$  (equal to 0.027) and the estimated depression  $T^s - T^{cs}$  of the coherent spinodal

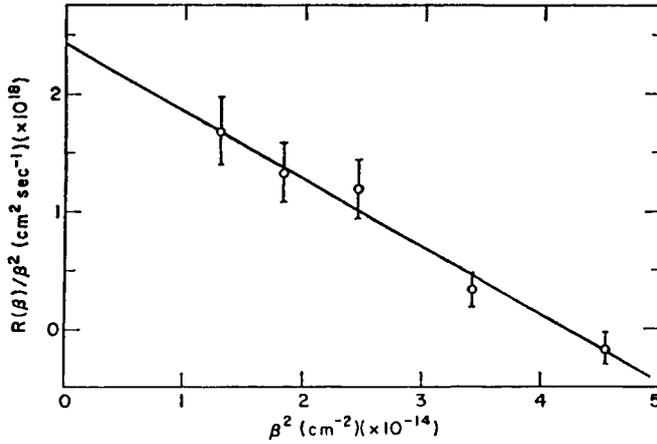


FIG. 16.6. Plot of  $R(\beta)/\beta^2$  versus  $\beta^2$  for an Al 22 at% Zn alloy annealed at 65 C (after Rundman and Hilliard, 1967).

below the chemical spinodal is only 35°C. Rundman and Hilliard estimated  $(\partial^2 g/\partial x_B^2)$  from thermodynamical data and, according to their figures, the intercept on the  $\beta=0$  axis is  $-1.15D_{\text{chem}}$  and is not sensitive to errors in the estimates. The derived value of  $D_{\text{chem}}$  is  $-2.6 \times 10^{-22} \text{m}^2 \text{s}^{-1}$ , in satisfactory agreement with a value of  $-5 \times 10^{-22} \text{m}^2 \text{s}^{-1}$  obtained from extrapolation of high temperature data. The corresponding value of the gradient energy coefficient derived from the slope of the plot was  $K=1.6 \times 10^{-10} \text{J m}^{-1}$ , but this is much more critically dependent on the estimate of the second derivative of the chemical free energy.

Although the Rundman Hilliard results appear to have been confirmed by a later study using an alloy quenched from the liquid phase (Agarwal and Herman, 1973), there are some difficulties in the simple interpretation given.

The considerable X-ray intensity which is present initially indicates some decomposition during the quench, and this presumably means that, in experiments on this alloy, the very early stages of the decomposition were not examined. In investigations on other spinodal systems, including glasses, the observations do not agree so well with the above predictions, especially at higher wave numbers where the plot of  $R(\beta)/\beta^2$  versus  $\beta^2$  becomes non-linear. Cook (1970) noted that the discrepancy is apparently most marked when the initial condition approximates most closely to a uniform distribution of solute, and he therefore ascribed the curvature to the effects of random fluctuations mentioned on p. 729. Cook showed that the effect of thermal fluctuations is to modify eqn. (72.20) to

$$I(\beta, t) = \{I(\beta, 0) - L(\beta)\} \exp[2R(\beta)t] + L(\beta) \quad (72.22)$$

where  $L(\beta)$  is an intensity contribution from the Brownian fluctuations. For a random binary solid solution the ratio of the intensity scattered by an atom and that scattered by

an electron is  $x_B(1 - x_B)(f_A - f_B)^2$  and if the scattering power of an atom is measured in terms of multiples of this average value (Laue units), Cook's expression for  $L(\beta)$  is

$$L(\beta) = kT/[x_B(1 - x_B)\{(\partial^2 g/\partial x_B^2) + 2\varepsilon^2 Y'v + 2Kv\beta^2\}] \quad (72.23)$$

where  $x_B = v_{c_0}$  is the overall atom fraction of the  $B$  component. At low values of  $\beta^2$  the difference between eqns. (72.22) and (72.20) is significant only in the very early stages of the decomposition when the amplitudes of the composition waves are small. However, at higher values of  $\beta$ , where  $R(\beta)$  is negative, the effects of the random fluctuation term will be more noticeable as the intensity given by eqn. (72.20) will not become larger than  $L(\beta)$ . Equation (72.22) predicts that there will be a particular value of  $\beta$ ,  $\beta'_c$ , for which the scattered X-ray intensity remains constant, given by the condition

$$I(\beta'_c, t) = I(\beta'_c, 0) = L(\beta'_c) \quad (72.24)$$

This wave number depends on the initial distribution and should now be identified with the observed cross-over point on experimental curves such as those of Fig. 16.5; it is such that  $\beta'_c > \beta_c$  of eqn. (42.17).

An experimental test of eqn. (72.23) was made by Acuña and Bonfiglioli (1974) who also worked with aluminium zinc alloys but chose an alloy with  $x_{Zn} = 0.15$  and a transformation temperature of  $-45$  C. They measured absolute X-ray intensities and this, in principle, allows the two quantities  $K$  and  $\partial^2 g/\partial x_B^2$  to be determined independently, whereas the original Rundman Hilliard analysis gives only the ratio of these two quantities. Acuña and Bonfiglioli first pointed out that their results were inconsistent with the straightforward Cahn theory; for example, neither their  $\{I(\beta, t)\}$  versus  $t$  plots nor the derived  $D(\beta)$  versus  $\beta^2$  plots were strictly linear and  $\beta'_c/\beta_m$  was appreciably larger than  $2\frac{1}{3}$ , thus indicating that the experimental  $\beta'_c$  should not be identified with  $\beta_c$  of eqn. (42.17). An alternative procedure is to calculate  $K$  from eqns. (72.23) and (72.24), using the experimental value of  $I(\beta'_c)$  and an estimate of  $\partial^2 g/\partial x_B^2$  obtained, as in Rundman and Hilliard's work, by extrapolation from high temperature equilibrium data.

This value of  $K$ , which was less than one-half of the value obtained from the initial linear part of the  $D(\beta)$  versus  $\beta^2$  plot of the Rundman-Hilliard analysis, was then used to calculate  $L(\beta)$  for all  $\beta$  from eqn. (72.23) and hence to calculate  $R(\beta)$  as a function of  $\beta$  from eqn. (72.22). The curve thus obtained had approximately the shape required by eqn. (42.15) but  $\beta_c/\beta_m$  was approximately two instead of  $2\frac{1}{3}$ . Thus the Cook modification of the theory is not adequate to remove the discrepancies between Cahn's analysis and the experimental results, and a possible reason is that the modification applies only to a continuum. Because the analysis of the X-ray scattering data is being extended towards relatively small wavelengths (high values of  $\beta$ ), Acuña and Bonfiglioli suggested that it may be necessary to take explicit account of the crystal lattice. For this purpose they used a formulation of the theory by Cook and De Fontaine (1969) which incorporates both the discrete lattice approach and the concept of thermal fluctuations. Actually, however, it appears from the results that the main reason for the discrepancy is probably an incorrect estimate of  $(\partial^2 g/\partial x_B^2)$ .

It was emphasized above that Cahn's theory applies to a continuum but is also appropriate to a crystal lattice provided that the important composition fluctuations have wavelengths appreciably larger than the lattice parameter. This condition is approximately satisfied for spinodal decomposition but not for continuous ordering reactions. A discrete lattice version of Cahn's three-dimensional diffusion equation was first given by Cook *et al.* (1969) and was extended to include the effects of coherency strain energy by Cook and de Fontaine (1969, 1971). An outline of this theory is now described in order to relate it to the continuum treatment.

Consider a binary solid solution in which the atom sites define a primitive Bravais cubic lattice, i.e. the crystal structure is simple cubic, f.c.c. or b.c.c. The atom positions are then given by eqn. (5.1) with appropriate integral or half-integral values of  $u_i$ . Any specified distribution of  $A$  and  $B$  atoms over sites of the crystal will gradually change with time and after a time interval which is long in comparison with the mean time for an atomic migration we may define the new distribution by the probability  $x_B(u)$  that the site  $u$  is occupied by a  $B$  atom. This probability, which is defined with respect to the initial distribution, replaces the atomic fraction  $x_B$  of the continuum theory and we may similarly define a discrete composition variable  $c_B(u) = v^{-1}x_B(u)$ . If the probability  $x_B(u)$  is the same for all sites in the crystal, the solid solution is random and homogenous and will have a certain free energy per atom. For a non-random distribution, one contribution to the overall free energy of a region is simply the sum  $\sum_u g(u)$  taken over all the lattice sites of a region, where  $g(u)$  is the free energy per atom of the random solution of atomic fraction  $x_B(u)$ . By analogy with the continuum treatment on p. 184, we must now add an energy term which depends on the spatial rate of change of  $x_B(u)$ . Let  $\mathbf{r}$  be one of the set of vectors joining any atom to a nearest neighbour. The gradient energy  $Kv(\nabla x)^2$  of the continuum theory is then replaced by the discrete sum  $(Kv/2a^2)\sum_r [x_B(u+r) - x_B(u)]^2$ , the factor of a half arising because there are twice as many neighbours as there are directions. It was mentioned on p. 724 as a more complex form than that given by Cahn. However, the differences are significant only for short wavelength fluctuations and, in the case of spinodal decomposition (but not of continuous ordering), Cahn's expression may still be used. Thus the free energy per atom of a region of the crystal containing  $N$  atoms may be written

$$g_{th} = N^{-1} \sum_u \{g(u) + v^3 \varepsilon^2 Y' [c(u) - c_0]^2 + (Kv^3/2a^2) \sum_r c [c(u+r) - c(u)]^2\} \quad (72.25)$$

which is the discrete equivalent of eqn. (42.8). When the wavelength is not sufficiently large for eqn. (72.25) to be valid, we must replace the modulus  $Y'$  by a wave-vector-dependent quantity  $Y'(\boldsymbol{\beta})$  which varies with both the direction and magnitude of  $\boldsymbol{\beta}$ .

The interpretation of diffraction data in the very early stages of precipitation is often ambiguous, and in recent years direct observations of the processes leading to precipitate formation have become possible by the use of the techniques of the field ion microscope and of the atom probe. As an example, Miller *et al.* (1995ab) have carried out a very comprehensive investigation of the early stages of spinodal decomposition in iron-chromium alloys. Figure 16.7 is an atomic reconstruction of an iron-45% chromium alloy aged for 500 h at 773 K. The segregation into chromium-rich and iron-rich regions can

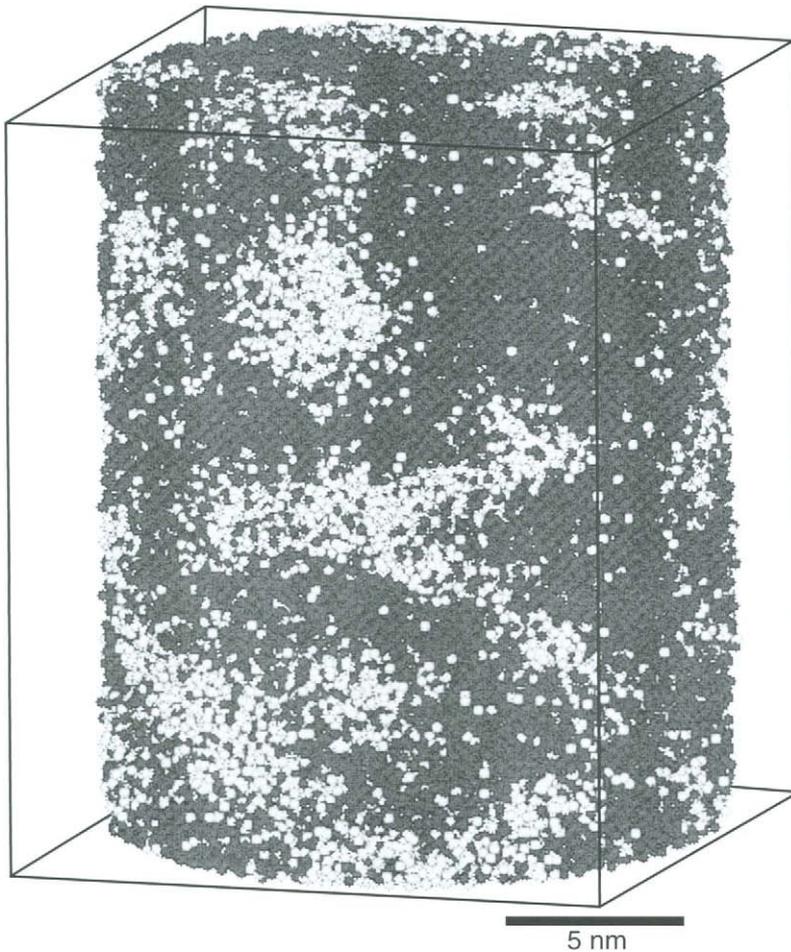


FIG. 16.7. Atomic reconstruction of an Fe 45 wt.% Cr alloy aged for 500 h at 773 K. The lighter spheres represent iron atoms and the darker spheres chromium atoms.

clearly be seen. Miller *et al.* used computer simulation based on the Monte Carlo algorithm and also obtained a numerical solution to the Cahn–Hilliard–Cook equation. Both these theoretical procedures yielded microstructures which were qualitatively similar to the experimental structures found by the atom probe technique. The atom probe has also been used to determine whether a particular transformation is nucleated or proceeds by spinodal decomposition.

### 73. CONTINUOUS PRECIPITATION

A supersaturated solid solution outside the spinodal region is metastable and all fluctuations will be resisted. The stable phase configuration can be attained only by a series

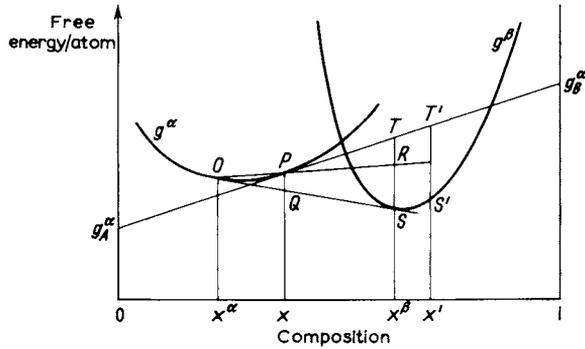


FIG. 16.8. Free energy versus composition curves for  $\alpha$  and  $\beta$  phases.

of favourable, but improbable, local fluctuations in structure and composition which eventually lead to the formation of stable nuclei, as already described in principle in Chapter 10. The only major difference is that the composition is now an additional free variable in the nucleation process.

Consider first the driving force for the formation of nuclei of varying composition; the argument is identical with that on p. 635. A supersaturated alloy of composition  $x$  (Fig. 16.8) will reduce the free energy from point  $P$  to point  $Q$  (on the common tangent to the two free energy curves) when precipitation is complete and the equilibrium structure has been attained, but the driving force for the initial nucleation per atom of the nucleus of composition  $x'$  is represented by the distance  $T'S'$  between the tangent to the  $\alpha$  free energy curve at  $x$  and the value of the  $\beta$  free energy curve at  $x'$ . The maximum driving force arises at composition  $x'$  where

$$(\partial g^\alpha / \partial x)_x = (\partial g^\beta / \partial x)_{x'} \quad (73.1)$$

and  $x'$  is necessarily greater than  $x^\beta$ , the equilibrium composition. As the slope  $PTT'$  increases as  $x$  increases, the maximum driving force will occur at increasing compositions as the supersaturation increases.

When the whole assembly is in the metastable  $\alpha$  phase, the free energy per atom is

$$g^\alpha = (1 - x)g_A^\alpha + xg_B^\alpha \quad (73.2)$$

where  $g_A^\alpha = (\partial G^\alpha / \partial N_A)$  is the chemical potential per atom of  $A$  in the  $\alpha$  phase. The corresponding equation for the free energy of the  $\beta$  phase of composition  $x'$  is

$$g^\beta = (1 - x')g_A^\beta + x'g_B^\beta \quad (73.3)$$

Suppose a nucleus forms in the metastable  $\alpha$  phase with composition  $x'$ . The (negative) free energy change per atom of this nucleus may be written  $\Delta g^{\alpha\beta}$  and is given by

$$\Delta g^{\alpha\beta}(x, x') = g^\beta(x') - g^\alpha(x) - (x' - x)(\partial g^\alpha / \partial x)_x \quad (73.4)$$

[see eqn. (65.2)]. The negative of this free energy change,  $\Delta g^{\beta\alpha}(x', x)$ , may be regarded as the driving force for starting the precipitation reaction; it reduces to  $(g^\alpha - g^\beta)$  for a polymorphic change in a pure metal or a compound. Substituting for  $g^\alpha$  and  $g^\beta$  gives

$$\Delta g^{\beta\alpha} = g_A^\alpha - g_A^\beta - x'(g_A^\alpha - g_B^\alpha + g_B^\beta - g_A^\beta)$$

as  $(\partial g^\alpha / \partial x)_x = g_B^\alpha - g_A^\alpha$  (see Fig. 16.8). The driving force has a maximum value at some composition  $x' > x^\beta$  and is then given by

$$\Delta g^{\beta\alpha}(x', x) = g_A^\alpha(x) - g_A^\beta(x') = g_A^\alpha - g_A^\beta \tag{73.5}$$

The composition  $x'$  is specified by the condition  $(g_A^\alpha - g_A^\beta) = (g_B^\alpha - g_B^\beta)$  or, in terms of the absolute activities,  $\lambda_A^\alpha / \lambda_A^\beta = \lambda_B^\alpha / \lambda_B^\beta$ . The extension of this relation

$$\lambda_A^\alpha / \lambda_A^\beta = \lambda_B^\alpha / \lambda_B^\beta = \lambda_C^\alpha / \lambda_C^\beta = \dots$$

gives the composition of maximum driving force for nucleation of the  $\beta$  phase in a multicomponent supersaturated  $\alpha$  solid solution.

According to classical nucleation theory, the rate of formation of precipitate nuclei of composition  $x'$  is given by Becker's equation (49.2) where

$$\Delta G_c = 4\sigma^3 \eta^3 / 27 \{ \Delta g^{\beta\alpha}(x, x') \}^2 \tag{73.6}$$

provided strain energy may be neglected.

There will be a value of  $\Delta G_c$  for each composition  $x'$  of the precipitate phase and, if a quasi-steady-state distribution of embryos of different sizes is set up for each composition, nuclei of various compositions will be in the process of formation. The composition giving the minimum  $\Delta G_c$  will obviously be preferred and nuclei of this composition will form so much more rapidly than every other composition that the others may be ignored. The calculation of the most favourable composition is not a simple matter, however,  $\sigma$  is also a function of  $x$  and  $x'$ . Only if  $\sigma$  is constant will the stable nucleus have the composition of maximum driving force.

An early approximation to the variation of the interfacial free energy with composition was suggested by Becker (1937, 1938). He used a simple nearest-neighbour model and, by counting bonds, he estimated the chemical part of the interfacial energy as

$$\sigma = \Xi(x - x')^2 \tag{73.7}$$

Becker assumed that the nucleus has the equilibrium composition  $x^\alpha$ , so that

$$\Delta G_c = 4\eta^3 \Xi(x^\beta - x^\alpha)^6 / 27\sigma^3 \{ \Delta g^{\beta\alpha}(x^\beta, x) \}^2 \tag{73.8}$$

However, if eqn. (73.7) represents a reasonable approximation to the interfacial free energy, the composition corresponding to the smallest  $\Delta G_c$  can be found; this was done by Hobstetter (1949) and Scheil (1952). Because of the strong dependence on the difference in composition which tends to make  $x'$  more nearly equal to  $x$  as the supersaturation increases, the effects of the surface energy and the maximum driving force oppose each other. Scheil showed that the variation can be quite strong; his calculations indicated that if the equilibrium compositions are 0.1 and 0.9, the composition of the initial precipitate

changes from 0.8 to 0.4 as that of the supersaturated solid solution increases from 0.14 to 0.22.

In Becker's theory, structural continuity across the interface is implicitly assumed so that the interfacial energy is obtained simply from the change in nearest-neighbour interactions. There is, however, an interfacial energy between solid phases of the same composition because of the disregistry of the two structures; this may be called structural interfacial energy, as distinct from the chemical interfacial energy calculated by Becker. If the boundary is semi-coherent, consisting of islands of fit separated by dislocations, it is possible to write

$$\sigma = \sigma_c + \sigma_s \quad (73.9)$$

where  $\sigma_c$  and  $\sigma_s$  are chemical and structural components respectively. The term  $\sigma_c$  varies rapidly with composition, [possibly as  $(x - x')^2$ ], but the variation of  $\sigma_s$  is due only to the slow change of lattice parameter with composition and it may either increase or decrease with increasing composition. The structural part of the interfacial energy is dependent mainly on the degree of disregistry and is given by an equation like (37.3). The energy should be similar to that of a grain boundary of similar dislocation content, although there is some evidence that the interfacial energy between  $\alpha$  and  $\beta$  crystals with high disregistry (incoherent boundaries) is rather smaller than the maximum grain boundary energy in either  $\alpha$  or  $\beta$  (Smith, 1952). For these incoherent boundaries, the separation (73.9) is not valid as it is no longer possible to identify a separate chemical term.

As already indicated in Chapter 8, the structural surface energy may be reduced if the nucleus is constrained into closer atomic fit with the matrix, but only at the expense of some strain energy which is then present in this configuration. In general, minimum energy will correspond to a semi-coherent nucleus of some definite disregistry less than that given by the free structures. Interfacial and strain energies will then be functions of the parameter  $\varepsilon_1$  (see p. 471) specifying the disregistry, and the total free energy change is

$$\Delta G = n\{\Delta g^{\alpha\beta}(x, x') + \Delta g_s\} + \eta\sigma(x, x')n^{2/3} \quad (73.10)$$

and at the saddle point of the free energy field

$$\begin{aligned} [\partial\Delta G/\partial n]_{y/R, x', \varepsilon_1} &= 0 \\ [\partial\Delta G/\partial(y/r)]_{n, x', \varepsilon_1} &= 0 \\ [\partial\Delta G/\partial x']_{y/r, n, \varepsilon_1} &= 0 \\ [\partial\Delta G/\partial\varepsilon_1]_{n, y/R, x'} &= 0 \end{aligned} \quad (73.11)$$

These conditions define the size ( $n$ ), shape ( $y/R$ ), degree of coherency ( $\varepsilon_1$ ) and composition ( $x'$ ) of the critical nucleus.

Consider now the formation of local clusters within the matrix by fluctuations in composition only. The energy increase associated with the formation of a cluster of given size and composition implies that in an equilibrium or quasi-steady state there will be a

statistical distribution of such clusters. As the energy increase associated with any cluster is  $n\Delta g(x, x')$  it follows that the density of clusters of any given composition  $x'$  is proportional to  $\exp(-n)$ ,  $n$  being the number of atoms in the cluster.

Suppose now that in Fig. 16.8  $x'$  is the composition of the most favourable nucleus as determined by eqns. (73.11). The increase in free energy required to form a segregate of this composition containing  $n$  atoms is thus  $n\Delta g^\alpha(x, x')$ . Leaving out the interfacial and strain energies, the change in free energy produced by transforming the cluster into a  $\beta$  nucleus is  $-n[\Delta g^\alpha(x, x') - \Delta g^{\alpha\beta}(x, x')]$  so that the energy for segregation does not enter into the expression for  $\Delta G_c$ , which is independent of the "path" of the reaction. Hardy and Heal (1954) pointed out, however, that the path is influenced by the relative magnitudes of  $\Delta G_c$  and  $n_c\Delta g^\alpha(x, x')$ . Figure 16.9 shows curves of  $\Delta G_c$  and  $n\Delta g^\alpha(x, x')$  for two different cases. In Fig. 16.9(a) less energy is required to form a segregate of size  $n_c$  than a  $\beta$  nucleus, and in a quasi-steady-state distribution there will be more segregates than  $\beta$  nuclei. The nuclei are likely to form directly from segregates as this process has a high probability, although segregates may grow to sizes larger than  $n_c$  before changing into nuclei. In the alternative case, Fig. 16.9(b) shows that there is a greater probability of embryos and nuclei of all sizes than of segregates, and the preferred path for formation of these embryos will not involve prior segregation within the parent phase. Obviously embryos and segregates of all compositions and sizes will continually be forming and disappearing again and it is very difficult to predict in detail the path of most probable nucleation. It is clear, for example, that clusters with compositions only slightly different from  $x$  will form more rapidly, but will turn into nuclei less readily because of the greater  $\Delta G$ . In some interpretations, precipitation begins with GP zone formation; in others, GP zones are formed only by spinodal decomposition.

Classical nucleation theory breaks down in the vicinity of a spinodal curve, not only inside the spinodal but also just outside it. As the spinodal is approached, the interface becomes very diffuse, the interfacial free energy becomes a gradient energy and tends towards zero, and the critical size, which according to classical theory decreases continuously, becomes very large and increases towards infinity. These changes were first

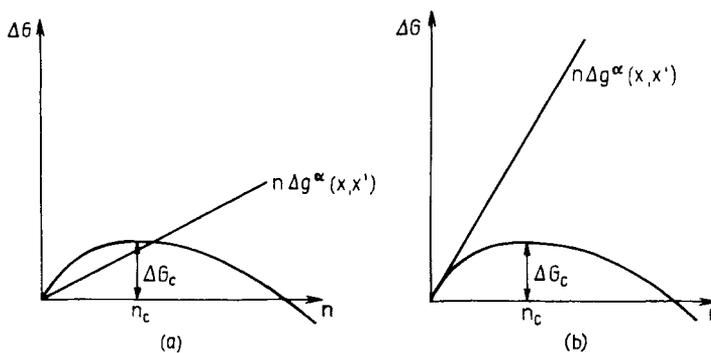


FIG. 16.9. Free energy required to form (a) a segregate before a nucleus and (b) a nucleus by direct fluctuations.

pointed out by Cahn and Hilliard (1958). The effect is sometimes called the “no theory” region because of the difficulty in giving a theoretical description of nucleation behaviour under such conditions.

Continuous or general precipitation is usually found in solid solutions of low supersaturation or with large misfit energies between the phases. These conditions imply that nucleation is difficult and continuous precipitation is the faster mode of reaction because of the difficulties in initiating “cell” growth. The low driving force also indicates that homogeneous thermal nucleation is unlikely and precipitation will usually begin from grain boundaries and dislocations.

The theory of the formal kinetics of continuous precipitation was outlined in Sections 54 and 58. If the rate of growth is governed by volume diffusion in the supersaturated  $\alpha$  phase, the linear dimensions of a precipitate particle are given by Zener’s equation (54.10) and the overall transformation law for nuclei existing at  $t=0$  is given by one of the equations (58.6), (58.19) and (58.21). Experimental tests of the assumption that the growth velocity depends on the coefficient of lattice diffusion were first made by Wert (1949) for the precipitation of carbide and nitride phases from solution in  $\alpha$  iron. Approximate verification of the assumption is possible by examining the growth rate of individual particles by an optical or electron microscope as, to an order of magnitude, eqn. (54.10) may be written

$$r^3 \sim (Dt)^{1/2} \quad (73.12)$$

for most reactions. Turnbull (1954) estimated  $D$  from this relation using published results on the precipitation of silicon,  $\text{CuAl}_2$  and  $\text{AgAl}_2$  from binary alloys of aluminium with silicon, copper and silver respectively. The results agreed with the experimentally determined lattice diffusion coefficients to within the rather large experimental uncertainty.

Wert’s studies on precipitation of carbides and nitrides from dilute  $\alpha$ -iron solutions were made by measurements of the internal friction peak associated with the stress-induced movement of the interstitial atoms. The height of this peak is proportional to the amount of solute  $\{c^m(t)\}$  remaining in solution. The results were analysed by plotting  $\log \ln\{c^m/c^\infty(t)\}$  against  $\log t$ , which is effectively equivalent to a plot of  $\log \log\{1/(1-\zeta)\}$  against  $\log t$ , as discussed in Chapter 12. Figure 16.10 shows the results for the precipitation of iron carbides, assumed by Wert to be  $\text{Fe}_3\text{C}$  at all temperatures although, in view of later work, this is unlikely. Metallographic work has shown that there are two stages of precipitation, the first of which, at temperatures below  $\sim 300^\circ\text{C}$ , corresponds to a carbide which is difficult to identify but is probably  $\epsilon$ -iron carbide. Cementite particles either nucleate directly on dislocations and grain boundaries or form from the  $\epsilon$ -carbide particles which redissolve at higher temperatures.

Experimental results for the precipitation of iron nitrides from iron–nitrogen solid solutions (Fig. 16.11) are less complete, but the general trends are similar. The precipitation is again in two stages, the first being the formation of an unknown nitride denoted “N phase 1”; this is probably the tetragonal phase later found by Jack (1951) to have composition  $\text{Fe}_{16}\text{N}_2$ . It may be seen that both the carbide and  $\text{Fe}_4\text{N}$  form in the early stages in accordance with eqn. (4.11), although the rate of precipitation of the nitride is

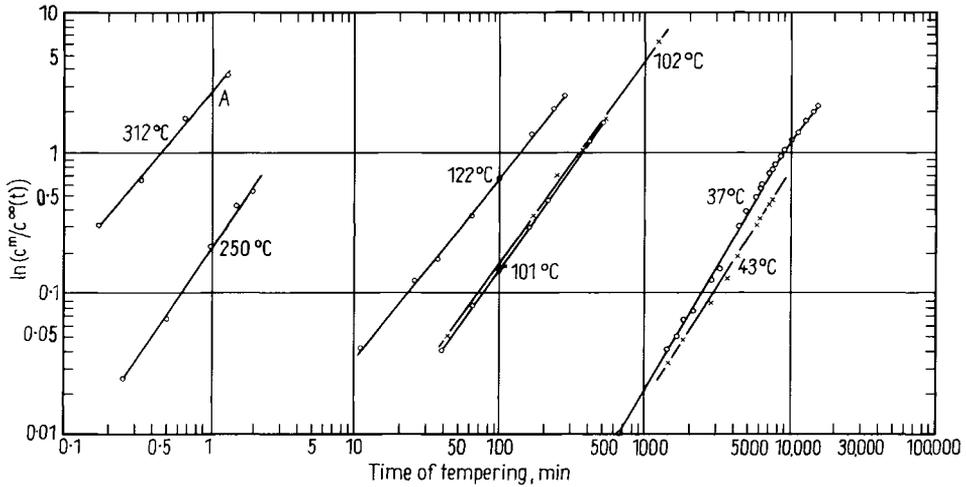


FIG. 16.10. Kinetics of the precipitation of carbides from dilute  $\alpha$ -iron solid solution (after Wert, 1949).

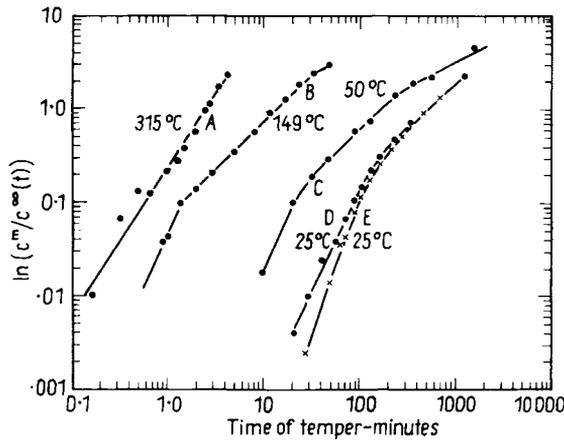


FIG. 16.11. Kinetics of the precipitation of nitrides from dilute  $\alpha$ -iron solid solution (after Wert, 1949).

too great for accurate comparison. The experimental value of  $n$  for iron carbide varies from 1.2 to 1.7 with a mean value of 1.45 which would correspond (see Table IX) to spherical growth with all nuclei initially present. Wert pointed out that, before impingement begins, the minimum value of  $n$  is 1.5 so that the overall transformation law suggests strongly that the rate of nucleation is zero. This conclusion is not substantiated by later work, however, and in fact various other types of growth could give  $n < 1.5$ . The measured  $n$  could thus include a contribution from a time-dependent

nucleation rate, and the real test of whether the nucleation rate is zero rests on metallographic observations and on the variation of transformation rate with temperature (see below).

Wert considered only the average value of  $n$  but later work has shown that the variation of  $n$  with temperature and with carbon content is a real effect. The most reliable results appear to be those of Pitsch and Lucke (1956) and Doremus (1960), and these are summarized in Table XIII.

These results all relate to unstrained samples of iron; strained samples are discussed separately below. At 0.022% C, the value of  $n$  increases slowly with decreasing temperature to about 60°C where there is an abrupt increase in  $n$ , followed by a further slow increase with falling temperature. Wert's results are very similar, although the carbon content of his specimens is not known accurately. At lower carbon contents, the value of  $n$  appears to be about 1.1 for all temperatures and concentrations examined. At fixed temperature, the precipitation rate increases rapidly with decreasing carbon content. The transition in the kinetic behaviour of the 0.022% C alloy and the absence of such a transition in the 0.012% C alloy was confirmed by Doremus by plots of  $\ln t_{1/2}$  against  $1/T$  as shown in Fig. 16.12.

The temperature variation of the precipitation rate depends on the number of nuclei initially present after quenching to any temperature, the nucleation rate at that temperature and the growth rate at that temperature. Wert found that the number of nuclei vary with temperature as the transformation rate at 50°C was greatly enhanced by first holding at 27°C to produce more nuclei. The following procedure was used to eliminate the variation of transformation rate with temperature caused by the variation in the initial number of nuclei. All specimens were given a preliminary quench to 27°C and

TABLE XIII. VALUES OF  $n$  IN AVRAMI EQUATION (4.11) FOR THE PRECIPITATION OF CARBON FROM ALPHA IRON

Wt. %C	Temperature (°C)	$n$	Reference
0.022	0	1.75	Doremus (1960)
	37	1.69	
	50	1.61	
	62	1.30	
	75	1.23	
	120	1.16	
	170	0.96	
0.013	37	1.1	Doremus
0.012	62	1.1	Doremus
0.017	90	1.19	Pitsch and Lucke (1956)
0.012	120	1.1	Doremus
0.015	120	1.08	Pitsch and Lucke
0.06	120	1.1	Doremus
0.08	120	1.1	Doremus
0.014	149	1.08	Pitsch and Lucke
0.016	170	1.02	Pitsch and Lucke

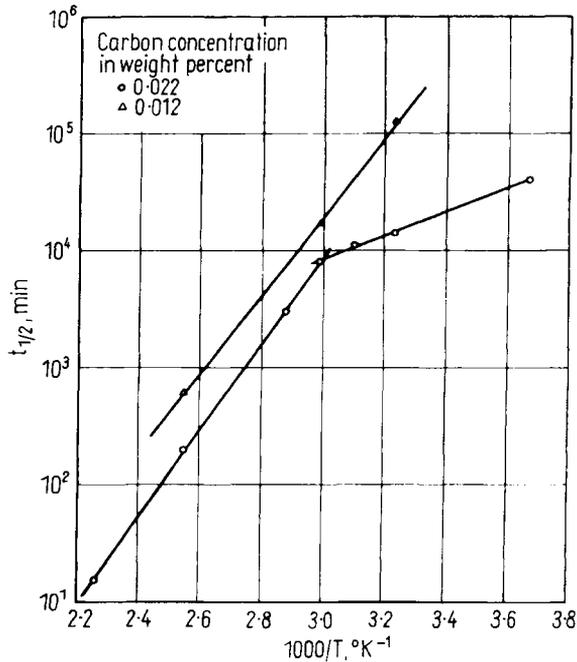


FIG. 16.12. Half-lives for carbide precipitation at different temperatures and initial concentrations of carbon atoms (after Doremus, 1960).

transformed to a fixed point before being transferred to the final transformation temperature. The time at this temperature to attain a fixed amount of transformation was measured and the results were found to obey a law

$$t_{\zeta} = C \exp(-\varepsilon/kT) \quad (73.13)$$

Wert obtained a value of  $\varepsilon = 73.5 \pm 7 \text{ kJ mole}^{-1}$  in the case of carbide precipitation. In a later survey of the measurements of the diffusion coefficient using elastic after-effect and internal friction methods below  $123^{\circ}\text{C}$  and the precipitation method at  $120$  and  $200^{\circ}\text{C}$ , Wert (1950) concluded that the activation energy for volume diffusion is  $84.3 \text{ kJ mole}^{-1}$ . The precipitation results were considered to be reasonably consistent with this energy as required by the growth theories. In a careful study, Doremus (1959) has verified, however, that the activation energy given by the thermal cycling method is appreciably lower than that given by the relaxation methods. Doremus could find no variation of activation energy with temperature of reaction, initial concentration of carbon and mechanical deformation up to 4% strain, the value obtained being  $71.4 \pm 2.1 \text{ kJ mole}^{-1}$ . In particular, the activation energy for growth does not apparently change below  $60^{\circ}\text{C}$  despite the marked change in kinetics as shown by Table XIII and Fig. 16.10.

The assumption that the experimental value of  $\varepsilon$  in eqn. (73.13) gives the activation energy for growth is only valid if all nuclei are pre-existing at the beginning of transformation. In Doremus' experiments, a slight variation of Wert's procedure was

used; all specimens were water-quenched and then held at temperature  $T_1$  until about 30–50% of the carbon had precipitated, and transformation was finally completed at  $T_2$  which could be either higher or lower than  $T_1$ . This proves directly that the number of particles precipitating is constant at both  $T_1$  and  $T_2$  and throughout the transformation (i.e. there is no nucleation) because any other assumption would require that the rate of nucleation has exactly the same temperature dependence as the growth rate. This kinetic conclusion has been verified by electron microscopy (Doremus and Koch, 1960; Leslie, 1961) except for the low temperature precipitation of high carbon alloys, which corresponds to the anomalous kinetic behaviour described above.

Accepting that the measured activation energy is that for growth, some reason must be found for the value being lower than that of lattice diffusion. The most obvious possibility is some form of diffusion short circuit, but this is not supported by experiments in which the density of dislocations or point defects was varied. There may be enhanced diffusion in the strain field of the precipitates although Doremus has reported that the activation energy is not changed by the size of the precipitate.

Next, consider the very complex problem of the overall transformation laws. Wert's original analysis assumed the validity of eqn. (4.11) for treating the impingement problem, and at small values of  $\zeta$  the theoretical curve gives a reasonable fit with the experimental results. In subsequent work, Wert and Zener (1950) paid especial attention to the later stages of precipitation and showed that eqn. (58.6) was in better agreement with the results than eqn. (58.9) (see Fig. 12.6). The two curves were arranged to coincide with each other and with the experimental points at  $\zeta = 0.6$ , in the later stages of the transformation, and eqn. (58.6) was then claimed to give good agreement up to 90–95% transformation.

These results would seem to have established the mode of precipitation of carbon in iron alloys, but later work has shown that the conclusion concerning the spherical shape of the particles was quite wrong. Pitsch (1955) claimed that a comparison of electrical resistivity and internal friction measurements showed that the carbide crystals form first as thin plates which subsequently thicken as growth continues. A similar conclusion was reached by Doremus (1957), who analysed the kinetic data using the assumption of constant concentration gradient at the surface of the growing particle. A good experimental fit is achieved with a curve obtained by numerical integration of eqn. (58.21) assuming a value of  $a = 99$ , and this also gives reasonable values of the parameters  $R$  and  $N^C$ . Doremus claimed that the limited experimental data available in 1957 fitted this equation much better than eqn. (58.6) used by Wert and Zener. When the later data of Table XIII are included, it remains true that the curve gives a good fit for all the kinetic observations on the precipitation of carbon except for the anomalous high carbon low temperature region. In view of the results from electron microscopy described below, this agreement must be largely fortuitous. The equation was used by Doremus in his thermal cycling experiments, but it does not invalidate the derivation of the experimental activation energy for the process as it may be treated purely as an empirical curve which correlates the data.

Various other proposals have been made to account for the experimental transformation laws, including the diffusion-controlled growth of particles of appreciable initial size and growth on impurity particles. Reference to Table XIII shows that the value of  $n$

between one and 1.5 could be explained in this way and a detailed discussion was given by Ham (1959). However, there are various difficulties with this hypothesis which does not agree with the metallographic evidence.

Disc-shaped carbide precipitates were first observed by Tsou *et al.* (1952) and by Pitsch (1957) using replica techniques, but the most informative results have come from transmission electron microscopy (Doremus and Koch, 1960; Leslie, 1961). This work shows that, under most conditions, the particles grow on dislocation lines or on sub-boundaries. The particles are plate-shaped rather than spherical (Fig. 16.13) and are often spaced at distances of less than 100 nm. This close spacing invalidates almost all the treatments of overall kinetics for diffusion-controlled growth. As the ageing temperature is lowered, the first precipitate changes from cementite with a {110} habit plane to (probably)  $\eta$ -carbide with a {100} habit plane. Doremus and Koch proposed a structure for this phase which is similar to that of martensite and virtually identical with that of  $\text{Fe}_{16}\text{N}_2$  which is the first phase to precipitate from solutions of nitrogen in iron at low temperatures. Some precipitate also forms within the grains in dislocation-free regions, even at temperatures as high as 300°C, and the number of such particles increases as the reaction temperature decreases. Leslie showed that the formation of particles in dislocation-free regions and the transition from cementite to the unknown carbide both take place at higher temperatures if the activity of the carbon in the solid solution is increased, e.g. by increasing the carbon content or by adding silicon.

As the anomalous behaviour of Fig. 16.10 is not associated with any change in the activation energy for growth measured by the thermal cycling method, it must result from a large increase in the number of precipitating particles. Electron microscopy shows that this is achieved by the change from predominantly dislocation growth at high temperatures to dislocation plus matrix growth at lower temperatures. Leslie's work shows that the abrupt change in kinetics is not due to the onset of general nucleation but to a change in the proportion of carbon precipitating at matrix sites. At higher temperatures, the carbides on dislocations grow at the expense of those in the matrix but, at lower temperatures, the matrix carbides grow to larger sizes than those on dislocations. This is illustrated in Fig. 16.13. Leslie also reports that the number of matrix precipitates is increasing with time at lower temperatures, i.e. some thermal nucleation is taking place. The density of matrix sites is about one in  $10^8$  iron atoms and the mean spacing between sites is about 100 nm. This means that the nuclei either form homogeneously or on small clusters of vacancies, there being insufficient vacancies to give dislocation loops from collapsed vacancy discs.

In the course of his experiments, Leslie made direct measurements of the rate of growth of carbide particles at 200°C where all particles nucleated on dislocations and had the cementite structure. He found the growth rate to vary as  $t^{-0.16}$ . The complexities of the whole of the growth process, as revealed by the electron microscope, are very great, and it seems very doubtful whether any generally applicable analytical expression can be derived. In addition to the points already mentioned, difficulties arise from such factors as the re-resolution of small particles and the coarsening of the whole structure in order to decrease the interfacial free energy, and the very inhomogeneous distribution of dislocations.

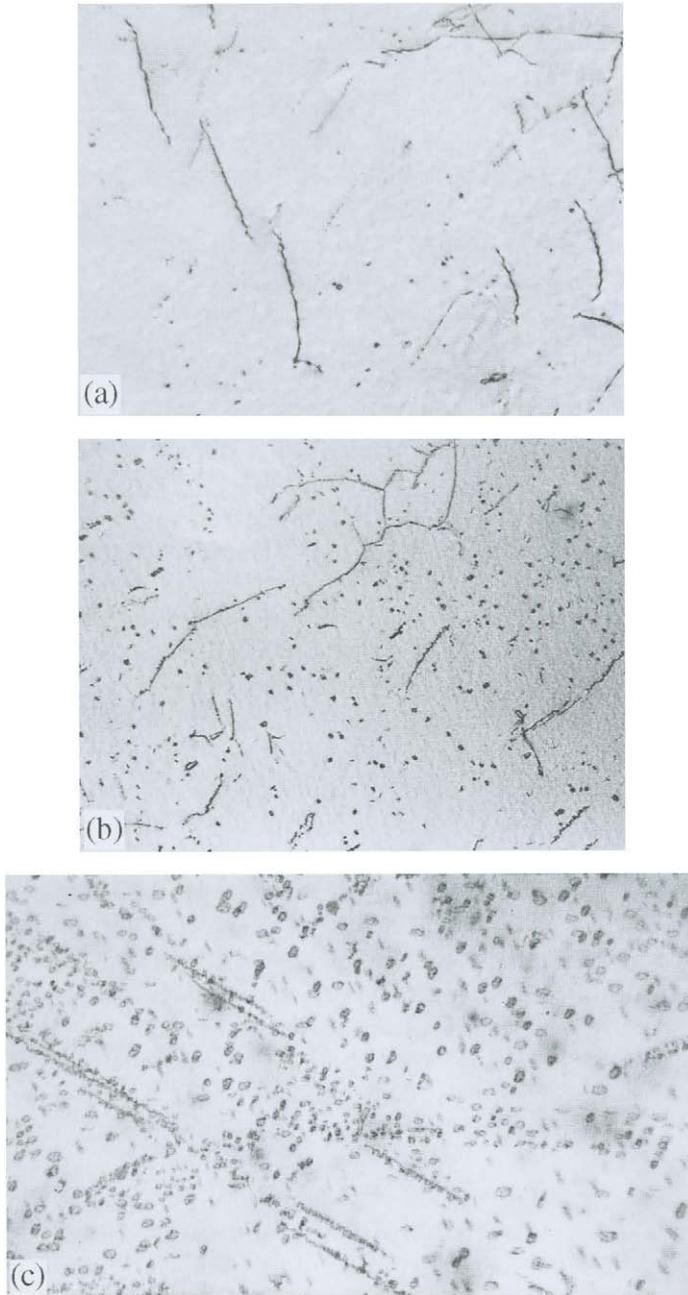


FIG. 16.13. Carbide precipitation in an iron-manganese-carbon alloy at 60°C (courtesy of Leslie, 1961): (a) after 500 min (original magnification  $\times 35\,000$ ); (b) after 64 h (original magnification  $\times 13\,500$ ); (c) after 7.2 days (original magnification  $\times 10\,500$ ).

Doremus has suggested that the value of  $n$  in the Avrami equation may be qualitatively attributed to the close spacing of particles along dislocation lines, and to their distribution. For widely spaced individual particles growing under diffusion control  $n = 1.5$ , for line sinks  $n = 1$  and for surface sinks  $n = \frac{1}{2}$  (see Table IX). Isolated dislocations correspond to line sinks if the particles are closely spaced along them and sub-boundaries or dislocation arrays produced by deformation may tend to become surface sinks.

Experimental data for the precipitation of nitrogen from iron are much less complete but the main features are probably very similar. The early work of Wert (see Fig. 16.11) shows three distinct stages, the first of which lasts up to about 20% of total precipitation and is represented by a value of  $n = 2.5$  in the Avrami equation. There is then an abrupt transition to  $n = 1.0$ , lasting to 80–90% transformation, and an ill-defined third stage. The high initial value of  $n$  was originally interpreted as indicating precipitation in the form of plates but this was based on erroneous conclusions about growth laws for diffusion-limited precipitation.

The above description refers to the precipitation of carbon from slightly impure iron. The corresponding behaviour of quenched steels is of great interest because of its commercial importance and also because the nature of the reaction is slightly different. In most precipitating assemblies, the parent phase is a supersaturated solution having the same symmetry and structure as one of the equilibrium phases into which it will eventually transform. Many quenched steels, however, have undergone a martensitic transformation, so that the parent phase has a different structure and symmetry, even though it is correctly regarded as a supersaturated solid solution. The low temperature precipitation of carbide from quenched steels comprises the process usually known as tempering.

The results of many workers using X-rays or optical or electron metallography indicated initially that in quenched plain carbon steels there are three distinct structural changes produced by low temperature ageing. These changes predominate in different temperature ranges. The lowest (0–160°C) is accompanied by an increase in hardness, whilst the second (230–280°C) gives slight softening and the highest (260–300°C) gives marked softening. The reactions are time-dependent and occur isothermally or on continuous heating; the stages are not entirely independent as the third overlaps the second and both may partly overlap the first.

The first stage was initially ascribed to the precipitation of a coherent non-equilibrium phase ( $\epsilon$ -iron carbide) with a h.c.p. structure together with the formation of a tetragonal low-carbon martensite ( $\sim 0.3\%$  C) at the expense of the primary martensite present in the quenched alloy. Early electron microscopy (Lement *et al.*, 1954) indicated that the  $\epsilon$ -carbide forms at sub-boundaries of the primary martensite. The second stage of tempering was believed to correspond to the transformation of regions of austenite retained in the structure after the quench, and the third to the formation of thin platelets of cementite from the  $\epsilon$ -carbide. These platelets change into spherical carbide particles, and the structure gradually coarsens on further tempering; electron microscope studies of this process were made by Hyman and Nutting (1956). Jack believed that there are two stages to the precipitation of cementite, which forms first as very thin platelets parallel to  $\{001\}$  cementite planes, such that the Laue conditions for diffraction normal to these planes are relaxed, and later as three-dimensional particles. Electron diffraction studies by

Calnan and Clewes (1952) and others, however, show that this is probably not true and that the plates are always three-dimensional.

Later work has shown that the sequence of reactions during tempering is much more complex than that given above. In a review of the later discoveries, Krauss (1984) distinguishes between “ageing” changes which occur within the initial martensitic phase and are essentially pre-precipitation phenomena and “tempering” changes which involve carbide precipitation. Considerable carbon atom rearrangement occurs even below room temperature and the first change at temperatures up to 100°C is now believed to be a clustering of carbon atoms within their original octahedral sites. The evidence for this comes from many techniques, but direct evidence has been obtained by atom probe microscopy (Miller *et al.*, 1983). A second ageing effect at temperatures below 100°C appears to be the formation of a modulated structure with carbon clusters separated at 1–2 nm on {102} planes (Nagakura *et al.*, 1983). X-ray line shifts indicate a gradual decrease in the tetragonality of the martensite, probably due to depletion of carbon between the clusters. Finally, in the range 60–80°C, a long period ordered structure develops with an orthorhombic structure and lattice parameters,  $2a$ ,  $2a$  and  $12c$  where  $a$  and  $c$  are the parameters of the martensite.

Precipitation of carbides takes place above 100°C, and begins with the formation of a transition carbide and the lowering of the carbon content of the surrounding matrix. This transition carbide as stated above was originally identified as  $\epsilon$ -carbide, but later work (Hirotsu and Nagakura, 1972) shows that it is in fact the orthorhombic  $\eta$ -carbide of composition (like  $\epsilon$ -carbide)  $\text{Fe}_2\text{C}$ . The precipitated particles are quite fine, about 2 nm in diameter, and arranged in rows. Because of the older work mentioned above, this is often referred to as the first stage of tempering.

At temperatures above 200°C retained austenite decomposes into ferrite and cementite. The measured activation energy for this process is  $\sim 115 \text{ kJ mol}^{-1}$ , consistent with that for diffusion of carbon in austenite. The austenite is initially present as thin layers between the martensite laths and decomposes to give relatively large cementite particles. This is commonly described as the second stage of tempering.

The third stage of tempering is the replacement of the transition carbide by the equilibrium carbide (cementite), and the decrease of the carbon content of the martensite to that of ferrite. These processes occur above 250°C concurrently with stage 2, but cementite is not always formed immediately; in high carbon martensites, the carbide which forms first is  $\chi$ -carbide and this is replaced by cementite only after prolonged tempering. Finally, in alloy steels, there may be a fourth stage of tempering at high temperatures in which alloy carbides are precipitated and give rise to secondary hardening.

The various tempering reactions are summarized in Table XIV reproduced from Krauss (1984).

Much research on steels concerns the decomposition of austenite and the formation of proeutectoid ferrite. Aaronson (1962; 1984) has emphasized the different morphological forms which precipitated ferrite crystals may have. In terms of a classification scheme due to Dubé and illustrated in Fig. 16.14 the main categories are grain boundary allotriomorphs, Widmanstätten side-plates and idiomorphs. Allotriomorphs have no

TABLE XIV. TEMPERING REACTIONS IN STEEL

Temperature Range (°C)	Reaction and Symbol (if designated)	Comments
-40 to 100	Clustering of 2 to 4 carbon atoms on octahedral sites of martensite; segregation of carbon atoms to dislocations and boundaries	Clustering is associated with diffuse spikes around fundamental electron diffraction spots of martensite
20 to 100	Modulated clusters of carbon atoms on (102) martensite planes (A2)	Identified by satellite spots around electron diffraction spots of martensite
60 to 80	Long period ordered phase with ordered carbon atoms arranged (A3)	Identified by superstructure spots in electron diffraction patterns
100 to 200	Precipitation of transition carbide as aligned 2 nm diameter particles (T1)	Recent work identifies carbides as eta (orthorhombic, Fe <sub>2</sub> C); Earlier studies identified the carbides as epsilon (hexagonal, Fe <sub>2.4</sub> C)
200 to 350	Transformation of retained austenite to ferrite and cementite (T2)	Associated with tempered martensite embrittlement in low and medium carbon steels
250 to 700	Formation of ferrite and cementite; eventual development of well spheroidized carbides in a matrix of equiaxed ferrite grains (T3)	This stage now appears to be initiated by chi-carbide formation in high carbon Fe-C alloys
500 to 700	Formation of alloy carbides in Cr, Mo, V, and W containing steels. The mix and composition of the carbides may change significantly with time (T4)	The alloy carbides produce secondary hardening and pronounced retardation of softening during tempering or long time service exposure around 500°C
350 to 550	Segregation and co-segregation of impurity and substitutional alloying elements	Responsible for temper embrittlement

external symmetry to reflect the symmetry of the atomic arrangement; in austenitic steels, they form on grain boundaries at a small driving force, i.e. at small supercooling below the phase diagram  $\gamma \rightarrow \gamma + \alpha$  boundary. These  $\alpha$  crystals grow most rapidly along the grain boundary, and less rapidly into the grains.

Several authors have attempted to measure nucleation rates of grain boundary allotriomorphs for comparison with theory. Using a sensitive etch, Langer and Aaronson (1979) were able to distinguish those ferrite crystals which had nucleated on grain edges from those which had nucleated on grain boundary surfaces. Considering only the latter, they computed the number density of nuclei per untransformed grain boundary area and found that they could get approximate agreement with theory only by assuming that the nucleus is of the pillbox type with coherent interfaces with both grains (Fig. 15.7), rather than the double convex lens type of Fig. 10.11. King and Bell (1975) also concluded that many ferrite allotriomorphs exhibit a low energy orientation relationship with both grains of the austenite and that faceting occurs even in the apparent absence of such relationships. It is indeed difficult to see how the ferrite manages to achieve a low energy, near-coherent interface with both, remembering that the austenite grains are supposedly

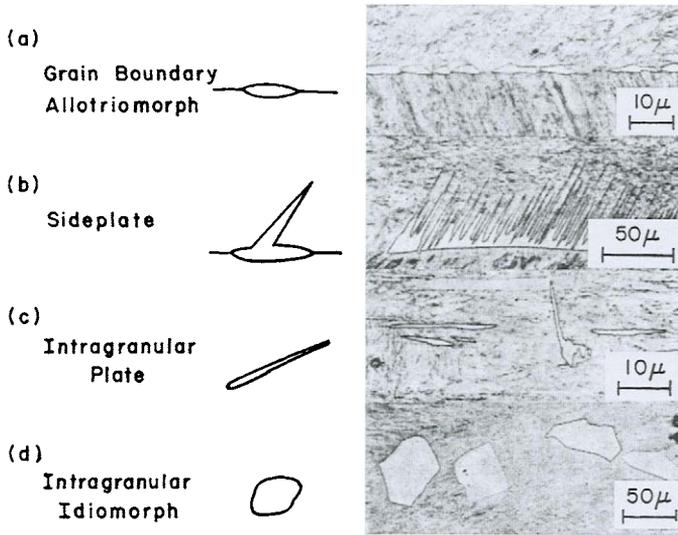


FIG. 16.14. Dubé morphological classification.

randomly orientated with respect to each other. Other measurements suggest that the precipitate has a Kurdjumov–Sachs orientation relation with one grain only, and a random orientation with respect to the other grain into which it grows.

Measurements of the growth rates (lengthening and thickening) of allotriomorphs present stereological difficulties, which have been overcome by special experimental techniques. Another complication is the suspected presence of a number of low energy facets mixed in with the generally incoherent sections of interface. Nevertheless it appears probable that the growth kinetics of such an averaged interface will not be greatly different from the assumption of a smooth (elliptical) interface which is completely disordered. Various experiments indicate that the thickening rate is proportional to  $t^{-1.2}$  as is expected for diffusion control. Purdy and Kirkaldy (1963) used decarburization and slow cooling into the austenite + ferrite region to establish essentially planar layers of polycrystalline ferrite at each end of a bar of an iron–carbon alloy. After annealing at a constant temperature in the austenite–ferrite region to eliminate composition gradients, the bar was isothermally reacted at a higher temperature to produce growth of the austenite relative to the ferrite. The growth was found to be parabolic and the rate constant agreed well with that calculated on the basis of a constant diffusion coefficient.

Thermionic emission microscopy was used by Kinsman and Aaronson (1973) to measure the thickening of a Fe–0.11% C alloy. The thickening rate was found to be parabolic, with a rate constant which showed considerable scatter but agreed approximately with the theoretical value of the rate constant  $\alpha_1$ , given by

$$\alpha_1 = \{2(c^\beta - c_\infty^\alpha)/(c^\beta - c^m)\}^{1/2} \quad (73.14)$$

[see eqns. (54.10) and (54.11)].

Measurements of both lengthening and thickening rates by Bradley *et al.* (1977) showed parabolic growth in three iron-carbon alloys, with carbon contents varying from 0.11% to 0.42%. A double spherical cap is considered to be the best approximation to the shape of the growing allotriomorphic precipitate, and the diffusion solution for an oblate ellipsoid was considered to be a reasonable approximation. The results were compared with predictions obtained by Atkinson's modification to the diffusion solution to allow for variation of the diffusion coefficient with composition (see p. 494). The ratio of the experimental to the theoretical growth rate was less than unity under all conditions investigated, and was especially low at low carbon concentrations and low supercooling. For the 0.11% alloy, the experimental rate constant is only 0.1 of the theoretical value at 840°C. The lower experimental values were again tentatively attributed to the development of low energy facets which slow down growth in their immediate neighbourhood.

The nucleation and growth of grain boundary allotriomorphs has also been studied in ternary Fe-C-X alloys. For nucleation, Enomoto and Aaronson (1985) again found the number density of nuclei formed on grain surfaces as a function of time. The results differed from those for binary Fe-C alloys inasmuch as a steady state was attained rather readily in all cases. The results were compared with classical nucleation theory, again assuming a pillbox type nucleus, and when replotted as nucleation rate versus driving force in order to eliminate effects due to the shift of the  $\gamma/(\alpha + \gamma)$  boundary in the equilibrium diagram, it appears that the binary Fe-C alloy gives the greatest nucleation rate, which is slightly depressed, at constant driving force, by Co, Si and Mo, and more severely depressed by Ni and Mn.

Growth theories for binary alloys generally assume local equilibrium at the interface giving diffusion-controlled growth, the deviation from this condition required to drive the interface process being negligible. However, conditions are very different in a ternary Fe-X-C alloy, where X is a substitutional solute with a diffusion coefficient about equal to that of iron whereas the diffusion coefficient of carbon is much larger. The growth problem is now that of balancing the flux of X with that of carbon and one extreme possibility is to make the flux of the substitutional solute zero so that any substitutional solute is not divided into two distinct phases of different compositions. This is described as para-equilibrium.

In a ternary system in which growth of the ferrite phase is controlled by diffusion it is possible, as in a binary alloy, to assume that the phases in contact at the interface are locally in equilibrium with each other; this requires that their compositions are given by the ends of a tie-line of the two-phase field in the equilibrium diagram. However, this tie-line need not pass through the overall composition of the alloy, but rather is selected to satisfy the flux balance equation

$$\alpha_X = [D_C/D_X]^{1/2} \alpha_C$$

where  $\alpha_X$  and  $\alpha_C$  are separate growth coefficients for X and for carbon. This can lead to two extremes. At low supersaturations, above the line AB in Fig. 16.15, the interface velocity is given by

$$\Gamma = \frac{1}{2} \alpha_X (D_X/t)^{1/2}$$

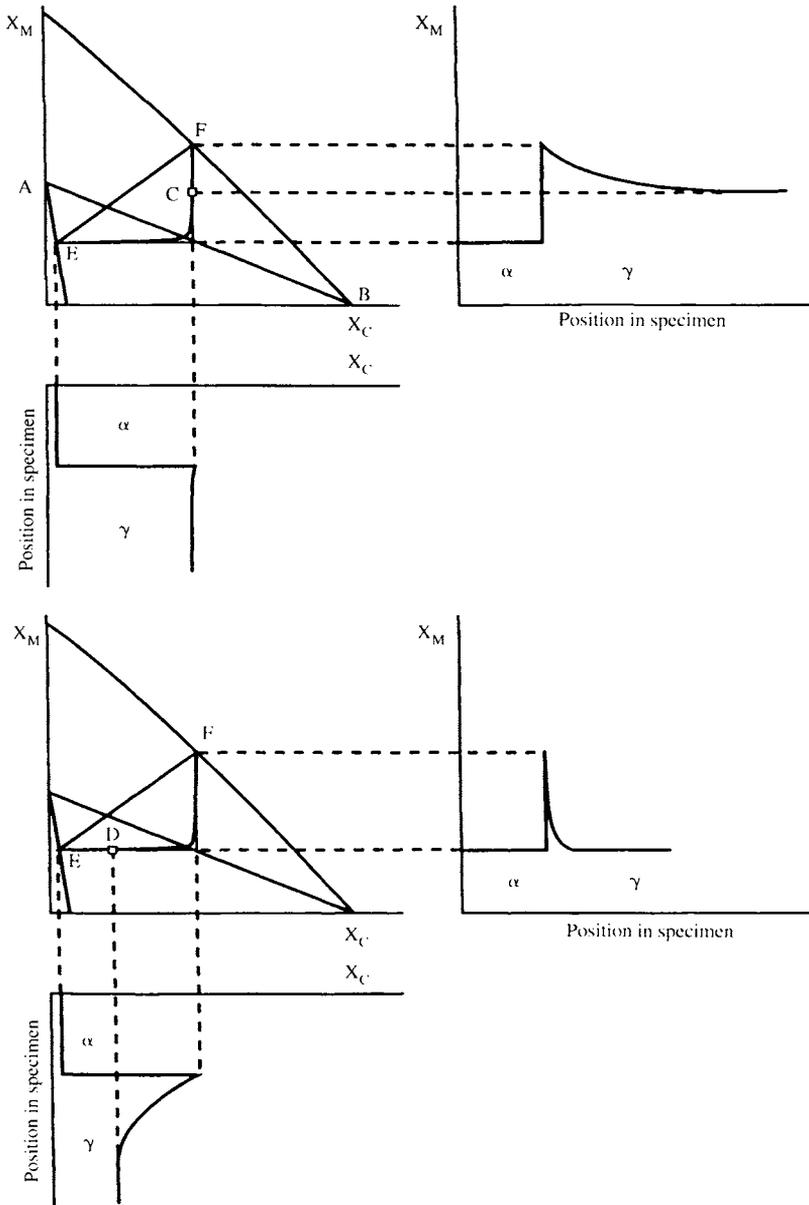


FIG. 16.15. Local equilibrium in a ternary system.

and it depends on the diffusion coefficient of X. Below the line AB, the growth velocity depends only on the diffusion coefficient of carbon and is given by

$$\Gamma = \frac{1}{2} \alpha_C (D_C/t)^{1/2}$$

In the first case, the carbon concentration gradient in the matrix is negligible, whereas in the second it extends far into the matrix. The X concentration gradient extends into the matrix in the first case but has only a very sharp and narrow “spike” in the second case. The growth velocity is naturally appreciably larger when controlled by carbon diffusion than by substitutional diffusion. The first case is usually referred to as local equilibrium–partition (LE–P) and the second case as LE–NP, the partition referred to being that of the substitutional solute. For an alloy of fixed composition, there may be a transition from LE–P to LE–NP as the temperature of reaction is reduced. For further details of the theory of local equilibrium in ternary systems see Coates (1972, 1973), Dehoff (1982) and Van der Ven and Delaey (1996).

Aaronson and Domian (1966) measured the partitioning of substitutional elements during the precipitation of ferrite from austenite in several Fe–C–X alloys. In most cases (Si, Mo, Co, Al, Cr and Cu), the supersaturation was rather high and there was negligible partitioning of the alloying element. For Mn, Ni and Pt alloys, however, partitioning was observed in each case above a certain temperature. An analysis of the results and comparison with the computed temperature of transition reveals some discrepancies (Reynolds *et al.*, 1984). The experimental parabolic rate constants for the thickening of grain boundary allotriomorphs in ternary iron alloys agreed with the computed values only for Fe–C–Ni alloys (Bradley *et al.*, 1982). The experimental growth rate was lower than expected for the para-equilibrium model for Fe–C–Mn and Fe–C–Cr alloys and higher than expected for Fe–C–Si alloys.

The above description has assumed disordered interfaces on which growth can occur at any site. However, in recent years much evidence has accumulated to suggest that, on coherent interfaces, growth generally occurs by a step mechanism. The steps are sometimes of atomic height and sometimes much larger. Steps of this kind have dislocation character in many situations and may glide along the interface to give a displacive transformation. However, the present consideration is directed towards the case where the steps or ledges grow by diffusion. This was not considered in Chapter 11.

Figure 16.16 shows the growth process as envisioned by Aaronson and his co-workers; this is sometimes described as the “terrace, ledge, kink” model. Atoms are absorbed into the new phase at the kinks in the steps and the velocity of the steps, and hence of the interface, is governed by diffusion in the matrix. The interface is considered to be either fully coherent or partly coherent and this implies that there is a lattice correspondence. The effect of the correspondence is to cause a shape change in the transformed region (see Fig. 16.19) and this is independent of whether or not the new phase has the same composition as the old phase. If the compositions are different, the rate of motion of the step is governed by diffusion in the matrix; a theory of ledge growth under diffusion control was first given by Jones and Trivedi (1971). The rate of advance of a surface swept by ledges is

$$\Gamma_s \approx h/\lambda u_s \quad (73.15)$$

where  $h$ ,  $\lambda$  and  $u_s$  are the ledge height, spacing and velocity respectively. The calculation of the ledge velocity is now done by considering the mass balance across the “riser” of the

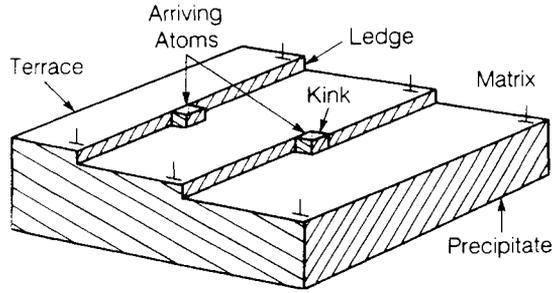


FIG. 16.16. Growth process envisioned by Aaronson and co-workers, described as the "terrace, ledge, kink" model.

ledge using concentration profiles obtained from the solution to the diffusion equation. A principal difficulty which arises is the specification of the boundary conditions at the step, where a uniform composition along the riser will not correspond to the actual situation. Analytic solutions to the diffusion problem were given by Jones and Trivedi (1971) and by Atkinson (1981) on the assumption, which has some experimental justification, that  $u_s$  is constant, and that the flux is constant over the riser surface. Isoconcentration contours, with respect to coordinates fixed in the moving interface, are shown in Figs. 16.17(a) and 16.17(b); the asymmetry in these contours is essentially opposite in the two treatments. Figure 16.18 shows the predicted Péclet number  $p = u_s h / 2D$  as a function of the supersaturation  $c_0 / (c^{\beta} - c^{\alpha})$ . An extension of both treatments (Jones and Trivedi, 1975; Atkinson, 1981) considers the growth of multiple intersecting and non-intersecting ledges.

Computer simulations have been used to investigate the growth of a stepped interface. Doherty and Cantor (1982) used a different boundary condition (zero supersaturation along the riser) and obtained very different results from those of the analytical treatments; for example, well separated steps never reached a steady state with a constant velocity. More recent simulations by Enomoto (1987) show that the lateral velocity of a single ledge agrees well with the Atkinson theory but not with the Jones and Trivedi predictions. The growth rate decreases with time and approaches a constant value. The growth rate normal to the terraces approximated to a parabolic growth law when it was due to an infinite train of ledges moving over the surface and, for closely spaced ledges, this growth rate could exceed that of a flat incoherent interface. However, the growth rate of the macroscopic interface formed by the terraces and ledges never exceeds that of a flat incoherent interface of the same orientation. For a finite chain of ledges, the velocity progressively decreases on going backwards down the chain, so that the chain broadens with time.

Transformations which require a change of composition but which are nevertheless accompanied by a systematic change of shape are apparently not uncommon although they were not considered in the initial division into displacive (martensitic) and reconstructive (nucleation and growth) types. The essential condition for a shape change (which must be an invariant plane strain) is that a lattice correspondence be maintained

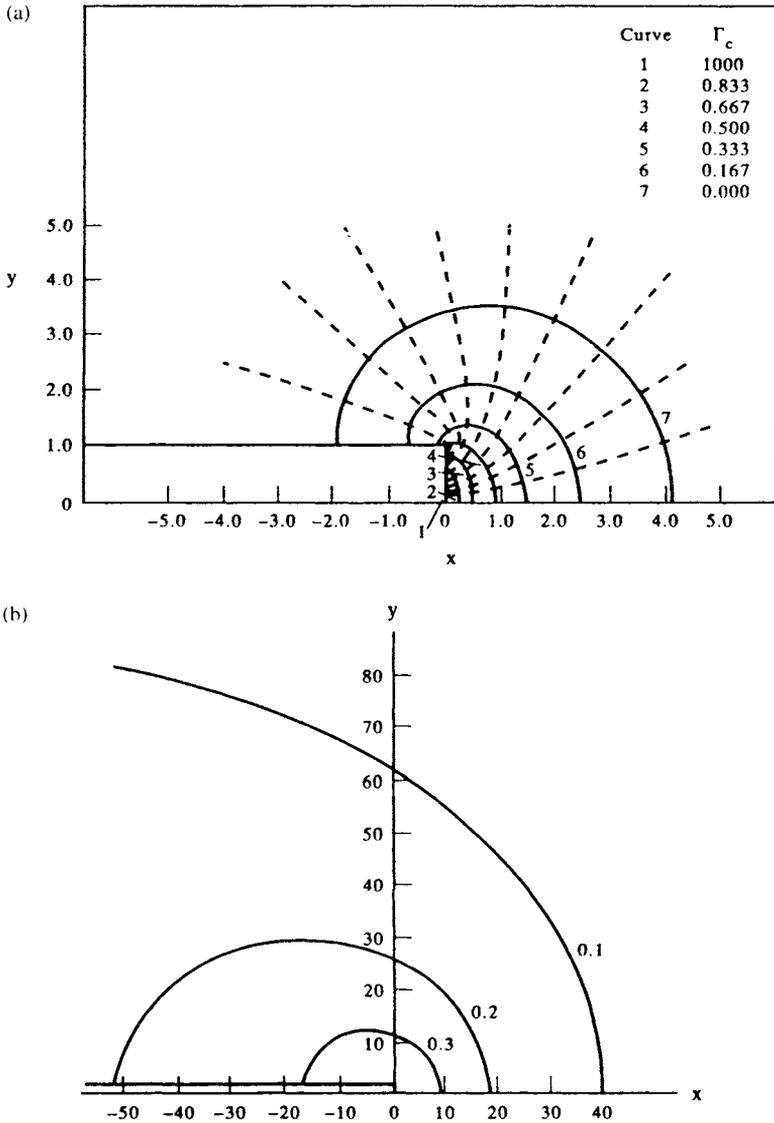


FIG. 16.17. Diffusion contours around a step in an interphase interface: (a) after Jones and Trivedi, 1971; (b) after Atkinson, 1981.

during the growth. Figure 16.19 illustrates a diffusional change in which coherency is maintained and which produces the shape change characteristic of a displacive reaction. Such a change has been called displacive-diffusional. Note that if new sites are created on one side of the plate or sites are eliminated on the other side, the macroscopic shape change effects will be eliminated.

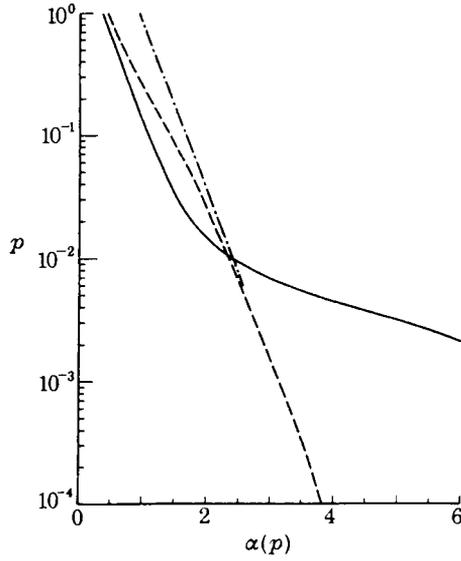


FIG. 16.18. Predicted Péclet number as a function of the supersaturation for ledge growth (after Atkinson, 1981).

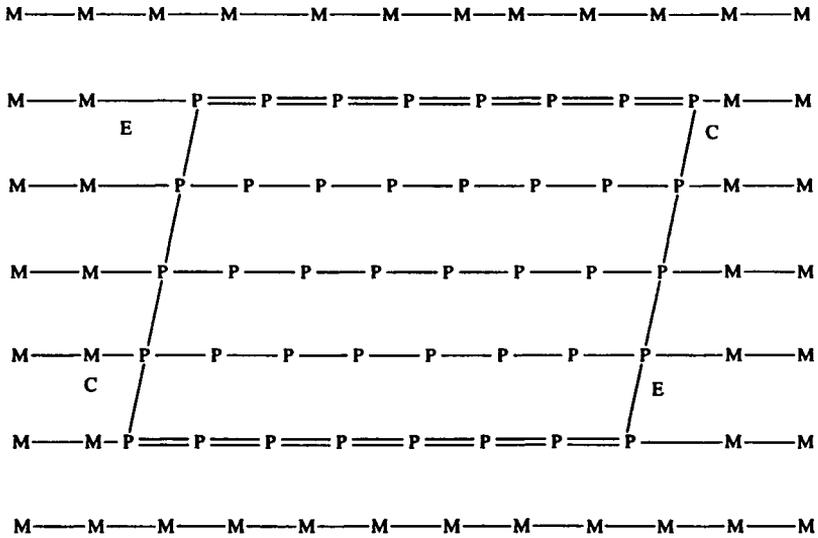


FIG. 16.19. A displacive diffusional transformation. The lattice correspondence across the coherent interface results in a shape deformation in the transformed region, but this could be eliminated if new sites were occupied at *E* (extended region) or old sites were abandoned at *C* (compressed region).

A well-known example is the formation of ferritic Widmanstätten plates in iron-carbon alloys and various steels at relatively high temperatures. Very careful measurements by Watson and McDougall (1973) show that the shape deformation is very large (0.34) but, together with the results for the habit plane and the orientation relations, it is in reasonably good agreement with the crystallographic theories of martensite. A shape change implies a lattice correspondence and an interface which is either fully coherent or semi-coherent (see Chapter 22). The crystallography is broadly consistent with the phenomenological theory of martensite (see Chapter 22), and this is surprising at first thought as the martensite interface is glissile and the necessity for diffusion prevents the interface migrating in a glissile manner. However, the interface can migrate by a diffusion-controlled step mechanism and its energy may be smaller than the alternative invariant plane type of interface which must contain misfit dislocations which have to climb as the interface migrates. Aaronson and his colleagues have laid great emphasis on the terrace-ledge-kink mechanism of growth. In the schematic Fig. 16.19 above and the Al-Ag alloys described below, the interface is rational; it is not so easy to understand how the correspondence is maintained in an irrational interface as in iron alloys.

The work on iron alloys has been described in some detail because of its commercial importance and because it illustrates the great complexity of overall transformation behaviour. Although many of the reactions treated have been low temperature changes, the greater mobility of interstitial atoms and the absence of many of the pre-precipitation effects found in substitutional alloys make it convenient to include these reactions in this section rather than in Section 75. There are, however, many non-ferrous examples of continuous precipitation, some of which will now be briefly considered.

The large misfit between parent and product lattices often encountered in continuous precipitation makes it improbable that the transformation will be homogeneously nucleated except at very large supercoolings below the thermodynamic transformation temperature. Experimental support for the conclusion that continuous precipitates are usually nucleated on lattice defects is very strong. Cold work almost always increases the rate of precipitation, although in most cases the observations do not establish whether this is due to enhanced nucleation or to more rapid growth, or both. However, many observations by optical or electron microscopy show that nuclei form preferentially on slip planes, sub-boundaries, grain boundaries and individual dislocations.

The precipitation of copper from supersaturated solid solution in germanium is an example of a continuous reaction which is extremely sensitive to structural imperfections and which can readily be followed by electrical methods. Logan (1955) found the rate of precipitation in crystals containing  $10^{10}$ - $10^{11}$  dislocations  $\text{m}^{-2}$  to be much greater than that in crystals with only  $10^8$  dislocations  $\text{m}^{-2}$ , and his work was extended by Tweet (1957). The advantage of a material like germanium is that the dislocation density can be controlled and measured rather accurately. Tweet found that the unprecipitated fraction of copper decreased exponentially with time, so that the isothermal transformation equation is of the form

$$\zeta = 1 - \exp(-kt) \quad (73.16)$$

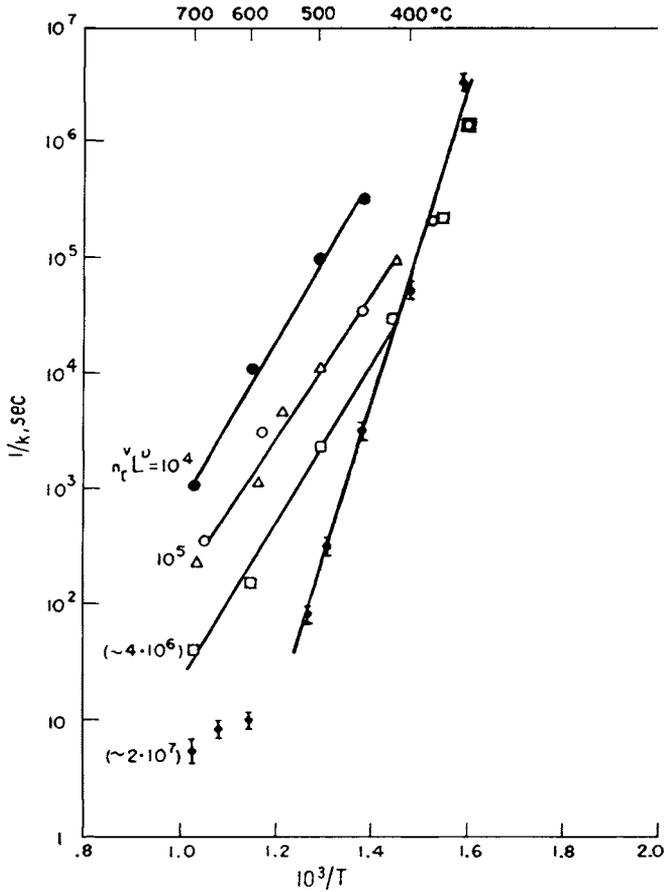


FIG. 16.20. Plot of the reciprocal rate constant  $1/k$  of eqn. (73.16) versus  $1/T$  for the precipitation of copper from germanium (after Tweet, 1957). The curves are labelled with dislocation densities, the smaller values being measured from etch-pit counts and the larger values (in parentheses) being estimated from the amount of deformation.

His results are shown in Fig. 16.20, where  $1/k$  is plotted against  $1/T$  for different dislocation densities. With a logarithmic scale for  $1/k$ , the curves are linear and parallel at high temperatures but merge into a single straight line of higher slope at sufficiently low temperatures for all dislocation densities greater than  $10^9$  lines  $m^{-2}$ . The experimental activation energies from this analysis are  $\sim 1.3$  eV at higher temperatures and  $\sim 2.7$  eV for the common line at lower temperatures.

Figure 16.20 shows that the rate of transformation at higher temperatures is greatly increased by increased density of dislocations. The rate-limiting factor is believed to be the diffusion of copper atoms to dislocations, which act as precipitation nuclei. This interpretation is supported by the value  $n = 1$  in the Avrami equation (see above) which is consistent (see Table IX) with diffusion-controlled radial growth of cylindrical

precipitates. The diffusion rate is very large and has been explained by Frank and Turnbull (1956) as being due to the presence of copper atoms in both substitutional and interstitial positions. The interstitial copper has a lower solubility by a factor of about 100, but a higher mobility, and in regions of high dislocation density interstitial atoms are responsible for most of the diffusion. For diffusion through a perfect lattice, the suggested mechanism is the dissociation of a substitutional copper atom, into an interstitial plus a vacancy, followed by the separate diffusion of the two defects. The vacancy diffuses more slowly than the interstitial so the observed diffusion coefficient is equal to the diffusion coefficient for vacancies multiplied by the fraction of the time that a vacancy is free from interstitials. Tweet found that the high temperature precipitation curves of Fig. 16.20 are consistent, both in magnitude and in temperature dependence, with the hypothesis that the coefficient of "dissociative diffusion" governed the precipitation. A similar mechanism is believed to govern the precipitation of copper from silicon where Dash (1956) has shown that the copper precipitates on dislocation lines, and may be used effectively to reveal the individual dislocations.

The dissociation of the substitutional copper is itself a thermally activated process and at sufficiently low temperatures it may determine the rate of the whole reaction. This is believed to explain the high activation energy and independence of the dislocation density observed at the lower temperatures. The time required for interstitial copper atoms and vacancies to move about in the lattice is now not critical and the distance they have to travel is not significant. The precipitate has, in effect, to wait for a sufficient supply of fast moving solute atoms to be formed from the almost immobile substitutional copper. The constants of the curve were shown by Tweet to be consistent with this theory and not to be consistent with any interpretation based on a slower diffusion process, which would require an impossibly high  $D_0$ .

The precipitation of silicon from aluminium-silicon solid solution is another continuous reaction which is very structure-sensitive. Because of the large misfit between the two structures, it is estimated that  $c^m/c^\alpha$  must be about 100 before homogeneous nucleation can be expected, whereas precipitation has actually been observed for  $c^m/c^\alpha \cong 5$ . A careful investigation of precipitation in a 1 wt.% Si alloy ( $c^m/c^\alpha \cong 10$  at 300°C) was made by Rosenbaum and Turnbull (1958). The kinetics of precipitation in the range 200–400°C were followed by measurements of electrical resistance and by metallographic methods. The specimens were first homogenized at 580°C and were then given various heat treatments before isothermal transformation; the results for transformation at 200°C were particularly striking. Direct quenching to 200°C or slow cooling to room temperature followed by up-quenching to 200°C both produced almost no precipitation and an intermediate stage of cold working at room temperature also had no effect. However, if the specimen was first quenched to room temperature and then up-quenched to the reaction temperature, rapid transformation ensued, and this transformation rate was further increased if the specimen was worked at room temperature before the up-quench. At temperatures above 200°C, precipitation occurred after all heat treatments but with similar large differences in rate. Microscopical examination of two specimens after reaction at 350°C showed large differences in the density of particles: these ranged from

$10^{20} \text{ m}^{-3}$  in specimens which had been water-quenched to room temperature before transformation to  $10^{12} \text{ m}^{-3}$  in air-cooled specimens. The results show that drastic cold work is ineffective in the transformation, but if adequate numbers of nuclei have been introduced by the water-quench treatment, cold work before transformation enhances the growth of the precipitate particles, presumably by providing diffusion short circuits.

In further experiments, it was found that the production of large numbers of nuclei ("inoculation") depends on ageing the specimen in a critical temperature range. Specimens directly quenched to  $-30^\circ\text{C}$  and then quenched immediately in liquid nitrogen, followed by fast up-quenching to the transformation temperature, were found to transform rather slowly. The fast rate was restored by pre-ageing such specimens in the temperature range  $-40^\circ\text{C}$  to room temperature; about 4 min at room temperature restored the rate to that found after a water quench. Figure 16.21 shows a set of results illustrating these effects.

The temperatures at which ageing is effective are roughly those at which point defects in aluminium can move at an appreciable rate, and also correspond to the formation of GP zones in some aluminium alloys. It is believed, however, that there is no appreciable clustering of silicon atoms in aluminium-silicon alloys prior to precipitation as no difference in the electrical resistivities of inoculated and non-inoculated specimens could be detected until after precipitation had begun. Rosenbaum and Turnbull attribute the phenomenon of inoculation to the presence of a supersaturation of vacancies immediately after the quench; in the slow-cooled specimens, this supersaturation would be removed during the cooling. During the ageing treatment, the defects reduce the supersaturation either by disappearing at dislocations, etc. or by the nucleation and growth of vacancy clusters. The motion of the vacancies necessary to achieve this could also lead to some clustering of the solute atoms.

Two possibilities for the nuclei are consistent with the experimental data. Formation of large numbers of solute atom clusters is improbable because of the failure to detect any change in electrical resistivity but, if only one atom in a thousand were in a cluster, this would still allow  $\sim 10^{21}$  clusters  $\text{m}^{-3}$ , each containing 1000 atoms. It is thus possible that clusters of this kind are the nuclei formed in the pre-ageing treatment; the supersaturation at room temperature is probably sufficient to give a reasonable rate of homogeneous nucleation, and the vacancies provide the diffusional flow to effect this nucleation. The nucleation and growth of clusters would then cease after a few minutes at room temperature, leaving most of the silicon in atomic solution, because the vacancy supersaturation had then become too low to provide an appreciable diffusional flow of silicon.

The other possibility is that aggregates of vacancies form the nuclei. It is impossible to decide whether such aggregates would form spherical voids or discs which collapse to form dislocation loops but the vacancy concentration ( $10^{25} \text{ m}^{-3}$ ) is adequate after a quench to provide  $10^{20} \text{ m}^{-3}$  of defects of either kind. A difficulty with the hypothesis of spherical voids is that they should tend to disappear after long-time ageing so that the inoculation effect should be destroyed; this was not observed. Also, it was found that the mechanical strength of the alloy markedly increased during the ageing process, and it is unlikely that this can be attributed to spherical voids. The formation of flat voids which collapse into dislocation loops overcomes both these difficulties, but it is not easy to understand why

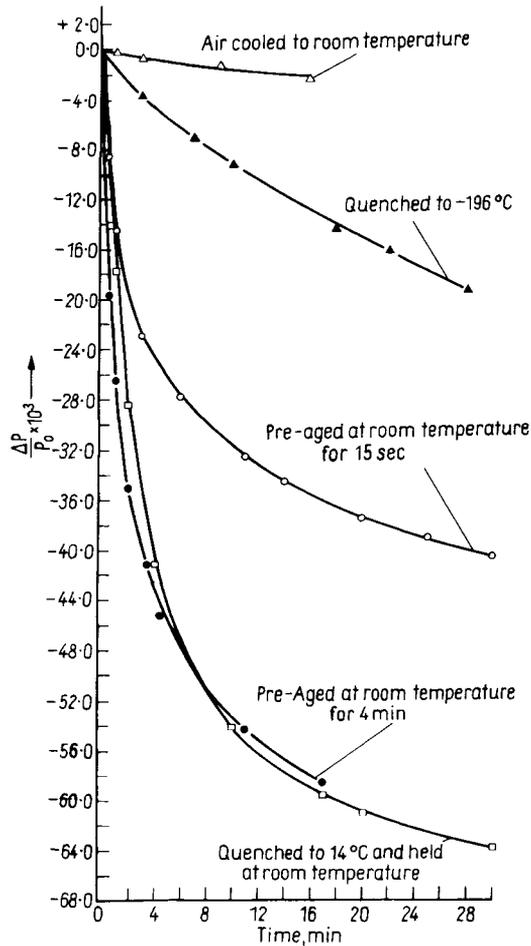


FIG. 16.21. The precipitation of silicon from aluminium silicon solid solution at 300 C as shown by the fractional change in electrical resistivity (after Rosenbaum and Turnbull, 1958). Specimens marked "quenched to  $-196^{\circ}\text{C}$ " and "pre-aged at room temperature" were quenched into iced brine and subsequently into liquid nitrogen before up-quenching to the ageing and/or transformation temperature.

such loops should be effective in nucleating silicon whilst the dislocations introduced by plastic deformation have no effect.

In the above, the importance of lattice defects in nucleation rather than in growth has been emphasized. Evidence that rapid diffusion along grain boundaries is important in precipitation at the boundaries has been obtained (Wood and Hellowell, 1961; Brandreth and Hellowell, 1962). The reaction studied was the precipitation of  $\alpha$ - from  $\beta$ -brass in the form of Widmanstätten side-plates. The results of metallographic and X-ray investigations indicate that whilst the density of potential nuclei may be sensitive to the misorientation across the boundary, the effective growth of a particle in the early stages of the reaction

depends on the migration of the grain boundaries. Widmanstätten side-plates almost always project only into  $\beta$  grains which are expanding and they have an orientation determined by the expanding grain. Growth seems always to begin by diffusion along the moving boundary to the triple  $\beta/\beta/\alpha$  junction but, beyond a critical size of a few microns, bulk diffusion becomes more important. The precipitate particles then grow backwards into the expanding grain and ultimately attain lengths exceeding the total boundary movement so that they extend on both sides of the original boundary position.

Finally, there are some non-ferrous examples of step growth, especially in the precipitation of  $\text{AlAg}_2$  from supersaturated Al-Ag solid solution. The  $\text{AlAg}_2$  phase is an ordered h.c.p. phase with alternate atomic  $\{0001\}$  planes consisting entirely of silver atoms and of equal numbers of aluminium and silver atoms. It forms a coherent interface with the f.c.c. solid solution in which the  $\{111\}_f$  plane of the solid solution is parallel to, and fully coherent with, the  $\{0001\}_h$  hexagonal plane. The cubic stacking can be changed into hexagonal stacking by the passage of a  $1/6\langle\bar{1}21\rangle$  Shockley partial dislocation through every other f.c.c.  $\{111\}$  plane; such a Shockley partial is equivalent to a step, two atom planes high, in the coherent interface (see Fig. 16.22). If the same Shockley partial passes through a series of alternate  $\{111\}$  planes, it produces a change of shape in the transformed region which is approximately a macroscopic shear<sup>†</sup> of magnitude  $8^{-1/2}$  but, if the shear direction is alternated among the three equivalent  $\langle 112 \rangle$  directions in a  $\{111\}$  plane, the macroscopic shear is reduced to zero (see Fig. 16.23). This is the situation in the martensitic transformation in cobalt and its alloys, described on p. 964, but in Al-Ag precipitation, the steps or Shockley partials cannot glide freely because of the need to change the composition of the hexagonal phase. Nevertheless, many investigations by Aaronson and his colleagues have shown that the growth mechanism does involve the motion of steps or ledges in the interface at a rate which is controlled by diffusion (see Fig. 16.16). Although diffusion-controlled, the mechanism is displacive in the sense that the macroscopic region may undergo a pronounced shape change. Experiments have shown that despite the possibility of eliminating the shape change as shown in Fig. 16.23 very often such a change is observed (Laird and Aaronson, 1967). All that is needed to produce a change of shape is a lattice correspondence which is sustained by the growth process, and it does not matter whether this is diffusional or diffusionless (see Fig. 16.19). With diffusional growth, the shape change could in principle be removed if atoms occupy sites not specified by the correspondence but this does not apparently happen, at least in Al-Ag alloys.

A step of multiple height, say  $n$  pairs of planes in the f.c.c. h.c.p. interface, may have a Burgers vector  $n$  times that of an ordinary Shockley dislocation, and hence will tend to dissociate again into elementary steps. However, its Burgers vector can be reduced by emission of lattice dislocations, combination with opposite lattice dislocations or, in this special case, by rotating the effective shear direction of successive elementary

<sup>†</sup>Strictly there is a small change in the spacing of the close-packed planes, so the shape change is an invariant plane strain rather than a simple shear. This volume change will be taken elastically at a step.

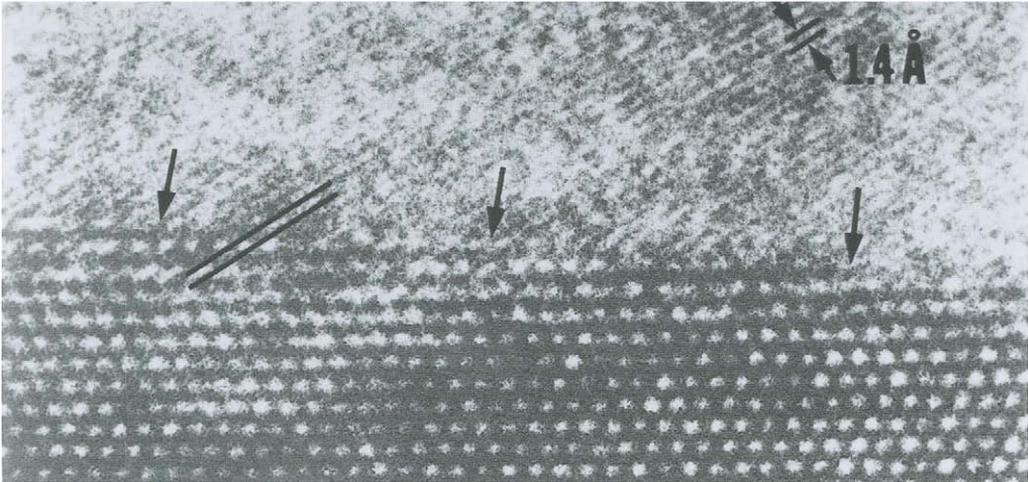


FIG. 16.22. Micrograph illustrating macroscopic shape change when all step vectors are parallel, and how averaging amongst the three possible vector displacements in one {111} plane will reduce the shape change almost to zero.

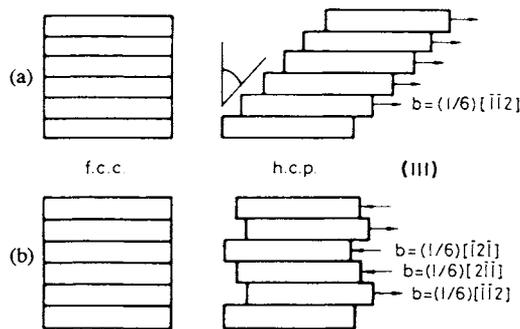


FIG. 16.23. Several steps (elementary dislocations) moving across the {111} face of an  $\text{Al}_2\text{Ag}$  precipitate in an Al 4.2 at.% Ag alloy (courtesy of Howe *et al.*, 1987).

components. This means that multiple height steps should not be uncommon in transformations of this type, and Fig. 16.24 shows an example from the Al–Ag system.

It may be helpful to summarize the different roles of lattice defects in continuous precipitation. It is well established (e.g. precipitation of copper from germanium) that dislocations act as nuclei, but this is not always the case; dislocations introduced via plastic deformation are sometimes ineffective in aiding transformation. Equally, much precipitation begins on grain boundaries, whether as allotriomorphs growing along the boundaries or as Widmanstätten side-plates. Finally, high concentrations of point defects seem to be effective in causing nucleation in some alloys, but the exact mechanism is not yet clear.

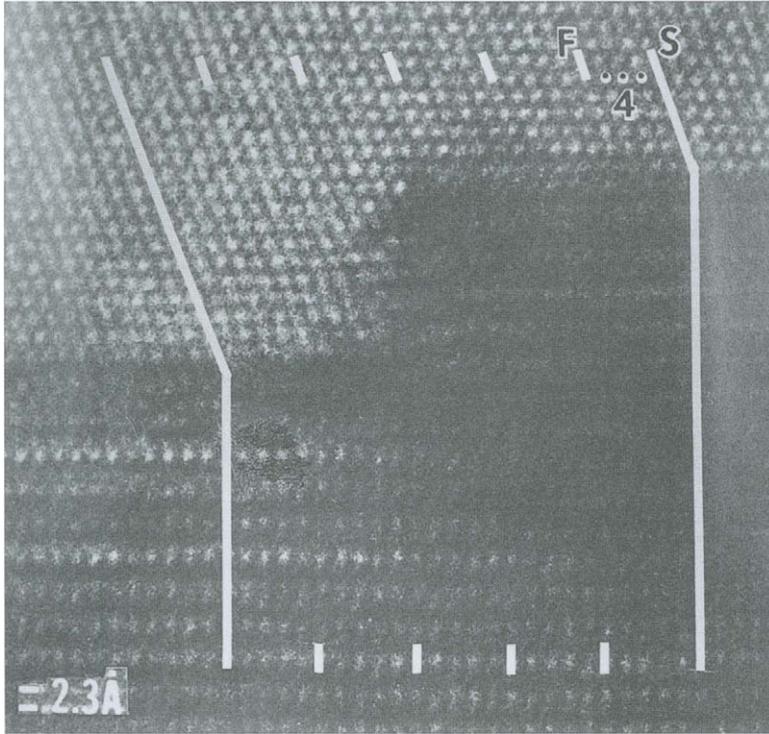


FIG. 16.24. Multi-height step on  $\{111\}$  face of an  $\text{AlAg}_2$  precipitate in an Al 4.2 at.% Ag alloy (courtesy of Howe *et al.*, 1987).

#### 74. DISCONTINUOUS PRECIPITATION

For many years, it was thought that discontinuous precipitation must be preceded by a stage of continuous precipitation or of clustering. Most workers no longer believe this to be true and in alloys such as lead–tin or gold–nickel the whole reaction appears to be discontinuous. The discontinuous mode of transformation requires the nucleation and growth of duplex cells of the product phases; these usually nucleate on a grain boundary and grow into one of the grains with a roughly hemispherical shape. The cell is most frequently in the form of lamellae of the  $\beta$  precipitate dispersed in the reorientated  $\alpha$  and, as a cell grows, the  $\beta$  plates must branch, or new plates must be nucleated, in order to maintain a constant interlamellar spacing. The growth of such a cell might be limited by the difficulty of nucleation or lamellation, but the available experimental evidence does not support this view.

The nucleation of a cell implies the formation of an incoherent boundary and hence is a difficult process in a region of good crystal. At a grain boundary, however, the necessary discontinuous interface is already present, and this provides a convincing reason why this type of precipitation almost invariably begins at grain boundaries. A plausible theory for

the origin of cell nuclei is due to C.S. Smith (1953). Suppose a  $\beta$  particle is nucleated in grain 1 at the junction of grains 1 and 2 of the  $\alpha$  phase. The orientation of this nucleus will probably be such as to minimize the energy of the interface between the  $\beta$  and grain 1. The  $\beta$  cannot grow readily into grain 1 because there are no diffusion short circuits available, but it can grow into grain 2 because there is an incoherent boundary between the supersaturated  $\alpha$  of grain 2 and the  $\alpha + \beta$  cell. As the cell grows, the  $\beta$  branches in order to keep the spacing constant and reorientation of the  $\alpha$  occurs at the cell boundary. According to this view, the growth of the cell is somewhat analogous to the encroachment of one grain on its neighbour during grain growth. The orientation of the  $\alpha$  in the cell should not be related in any way to that of the grain into which it is growing, but it should be nearly identical with that of the supersaturated  $\alpha$  on the other side of the boundary (i.e. grain 1) from which the cell originated. Smith published two striking micrographs to support this view. The first of these was a standard optical micrograph of the cellular structure in a partially transformed zinc-copper alloy containing 2% copper, whilst the second micrograph showed the same area under polarized light, in order to reveal orientation differences. Under the polarized light, the major phase in the cell could be seen to have an orientation which was quite different from the grain into which it was growing but which was indistinguishable from that of the grain into which it was not growing.

The Smith mechanism is not the only means of discontinuous nucleation. Cells sometimes form within the grains, presumably on dislocations or sub-boundaries, and on boundaries cells sometimes grow alternately into one or the other grain. Discontinuous precipitation may be regarded as involving both the formation of a precipitate and the displacement of the grain boundary. Butler and Williams (1982) divided suggested mechanisms into those in which the precipitate causes the boundary motion and those in which a moving boundary initiates precipitation. Figure 16.25 shows schematically how a moving boundary might initiate a transformation. The boundary need not move into the same grain along its whole length; Fig. 16.26 shows the "S-mechanism" which depends on the bulging outwards of the boundary on the convex side of the local curvature. Butler and Williams state that it is also possible for growth to occur on the locally concave side.

The detailed theory of cellular growth has already been described in Section 55 where it was mentioned that experimental work on the kinetics of discontinuous reactions supports the conclusion that cell boundary diffusion is often of great importance. The main evidence for this assertion comes from the very low temperatures at which discontinuous precipitation has been observed; lead-tin alloys will transform at carbon dioxide temperatures ( $-78\text{ C}$ ) and gold-nickel and gold-cobalt alloys transform slowly at room temperature. In all these examples, the activation energy for volume diffusion, as deduced from high temperature measurements, would give a negligible reaction rate at the temperatures mentioned so the conclusion that some other easier diffusion process is operative seems unavoidable.

The most widely studied example of discontinuous precipitation is probably the precipitation of tin from lead-tin solid solution. The rate of precipitation was measured by Borelius and his co-workers (Borelius *et al.*, 1944; Borelius and Säfsten, 1948; Nystrom, 1949) and by Turnbull and Treafis (1955). The kinetic work led to the

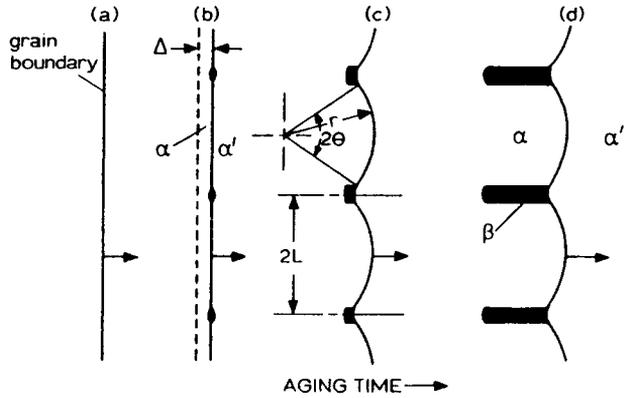


FIG. 16.25. Model of initiation of discontinuous precipitate (after Fournelle and Clark, 1972).

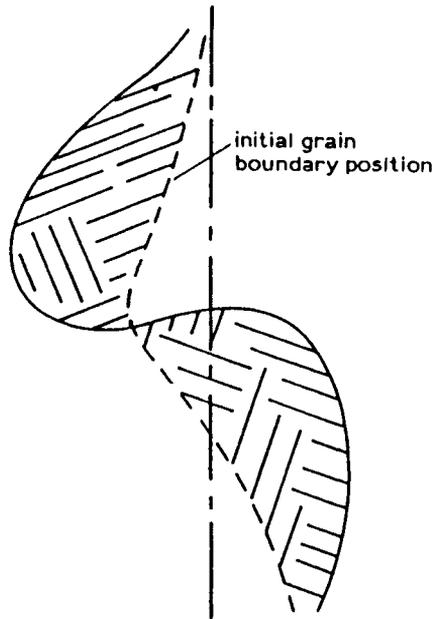


FIG. 16.26. Discontinuous precipitate forming on both sides of a grain boundary, depending on the curvature of the boundary (after Williams and Butler, 1981).

conclusion that the reaction involves the nucleation and growth of duplex cells, and this was confirmed by metallographic examination. Tiedma and Burgers (1954) also found that the discontinuous precipitate could form at room temperature in alloy single crystals, presumably after nucleating on an interior dislocation or dislocation group. Turnbull and

Treafitis found that the whole precipitation reaction occurs in two stages, as had been suggested by some of the earlier workers.

Stage 2 represents the discontinuous coarsening of the stage 1 structure and the completion of the precipitation process. The kinetic law applicable to the first process was found to be

$$\zeta = 1 - \exp(-kt^3) \quad (74.1)$$

which is consistent with hemispherical growth if all nuclei are present initially (see Table IX). The same kinetic law was found by Nystrom (1949) using calorimetric measurements.

In eqn. (74.1)  $\zeta$  refers not to the amount of transformation as previously defined but to the volume fraction of the transformation achieved in stage 1. Thus  $\zeta = 1$  represents the completion of the fast reaction, but the experimental results show that at this stage only about 60% of the excess solute has been removed from the supersaturated  $\alpha$  solution. This means either that the first stage ends before the cells have occupied the whole volume of the specimen, or that the duplex cells contain phases of non-equilibrium composition. As discussed in Section 55, a non-equilibrium segregation is predicted in Cahn's theory of cellular growth, and the experimental results imply that at room temperature  $Q^f \cong 0.6$  in lead-tin alloys. The authors also state that the fraction of excess solute precipitated in stage 1 does not vary with temperature, but the data do not seem to be sufficiently extensive to establish this contention.

The fast reaction is succeeded by a much slower stage 2 during which the remaining excess solute is precipitated at a rate about 100 times slower than the stage 1 reaction. Turnbull (1955) considered possible mechanisms for the two processes and concluded that incomplete segregation is the most probable explanation of the first stage. As there is ample evidence that the fast reaction involves diffusion along the cell boundary as the main segregation process, the slow reaction might correspond to diffusion through the lattice or via stationary dislocation pipes. However, Borelius and Larsson (1956) showed that the apparent activation energies of the fast and slow reactions are similar ( $\sim 170 \text{ kJ g atom}^{-1}$ ) and this was confirmed in the microstructural work of Turnbull and Treafitis (1958), who also showed that complete edge impingement of the cells occurred before the centres of the austenite grains had been transformed. It is thus possible that the segregation achieved during the fast reaction exceeds 60% of the equilibrium segregation, and the first part of the slow reaction corresponds to the migration of the cell boundaries after complete edge impingement. The much slower rate is presumably to be attributed not only to the reduction from  $n=3$  to  $n=1$  in eqn. (4.11) but also to a smaller value of  $\Gamma$  when edgewise growth is no longer possible.

The form of the kinetic law and the observation that the cells grow hemispherically show that the parameter  $k$  in eqn. (74.1) is given by

$$k = (2\pi/3)^n N^c \Gamma^3 \quad (74.2)$$

where  ${}^nN^c$  is the number of cell nuclei per unit volume. Turnbull and Treafitis measured the rate of precipitation in thermal cycling experiments in which the temperature of transformation was changed upwards or downwards after a certain percentage of the

assembly had transformed. From an analysis of these results, they showed that, for constant  $x^m$ , the number of nuclei increases with decreasing temperature until a critical temperature  $T^n$  is reached, below which  $N^C$  remains constant. The variation of  $\ln \Gamma$  with  $1/T$  is not linear, except possibly below  $T=273$  K.  $\Gamma$  increases with increasing temperature to a maximum value at a supersaturation of  $i \sim 5-6$ , and then decreases sharply again. Even for temperatures less than  $T^n$ , the rate of precipitation is very dependent on  $x^m$ . These and other deductions from the resistance measurements were subsequently confirmed by microscopical observations.

The temperature  $T^n$  corresponds to the "limit of retardation" in the work of Borelius and his colleagues and was interpreted by them as the spinodal temperature, below which the "thermodynamic barrier to nucleation" disappears. Turnbull pointed out, however, that the form of the kinetic reaction law is independent of whether precipitation is carried out above or below  $T^n$ , and that the cell nuclei appear to form on structural imperfections, and do not form homogeneously as the spinodal theory requires. The constancy of  $N^C$  below  $T^n$  thus seems to imply only that nucleation is no longer possible and that all available nucleation sites have been activated already. The conclusion that nucleation is difficult is strongly supported by some experiments of de Sorbo and Turnbull (1956) in which a calorimetric technique was used to follow the transformation in small powder samples of lead-tin alloys. Very slow transformation rates were obtained and were attributed to the absence of grain boundaries or other nucleating imperfections, a large proportion of the particles being single crystals. The precipitation rates were increased considerably by cold working and decreased to negligible values by prolonged homogenization.

As already noted, the observed precipitation rates cannot be explained by the hypothesis of volume diffusion. Typical experimental values of  $\Gamma$  and  $y^{\alpha\beta}$  substituted into eqn. (55.14) would require  $D \sim 10^{-14} \text{ m}^2 \text{ s}^{-1}$  at 300 K, which is  $10^8$  times larger than the diffusion coefficient extrapolated from results at higher temperatures and with lower solute contents. The corresponding value of  $D^B$  from eqn. (55.20) is  $10^{-10}-10^{-12} \text{ m}^2 \text{ s}^{-1}$ , and the probable activation energy for cell boundary diffusion is about  $38 \text{ kJ g atom}^{-1}$ , in place of the experimental result for volume diffusion of  $110 \text{ kJ g atom}^{-1}$  (Seith and Laird, 1932). Attempts to explain experimental results on the basis of experimental inaccuracy and unusual composition variation in  $D$  are unconvincing and Turnbull (1954) pointed out that when precipitated tin redissolves at high temperatures, the kinetics are satisfactorily described by an activation energy of  $120 \text{ kJ g atom}^{-1}$ .

The variation of interlamellar spacing with temperature and its relation to Zener's theory has led to some confusion in the literature. For a given  $x^m$ , the spacing is a function only of the reaction temperature, as illustrated by Cahn's experiments which prove that the new spacing is established immediately after either an upward or a downward change in temperature. Turnbull and Treafits analysed the measured variation of spacing with temperature in terms of the Zener prediction that  $y^{\alpha\beta}$  is proportional to  $1/\Delta g^1$  [eqn. (55.1)]. They wrote  $\Delta g^1$  as  $kT \ln i$  for a dilute phase where the supersaturation ratio  $i = c^m/c^\alpha$  and obtained a straight line plot of  $y^{\alpha\beta}$  against  $1/(T \ln i)$ . From the slope of the line, they concluded that  $y^{\alpha\beta}$  is about 100 times larger than the predicted spacing using a reasonable value of  $130 \text{ mJ m}^{-2}$  for  $\sigma^{\alpha\beta}$ . Cahn (1959) pointed out that this analysis is faulty, in as

much as  $kT \ln i$  is equal to the change in the chemical potential and not to the change in free energy  $\Delta g^l$ .

In a sufficiently dilute solution, the free energy equation is of the form (22.14), and the negative of the driving force is

$$\Delta g^l = kT[x^m \ln(x^\alpha/x^m) + (1 - x^m) \ln\{(1 - x^\alpha)/(1 - x^m)\}] \quad (74.3)$$

Using this equation in conjunction with the experimental spacing, Cahn calculated values of  $-y^{\alpha\beta} \Delta g^l / 2v$ , which is equal to  $\sigma^{\alpha\beta} / R^l$  [eqn. (55.31)]. These values range from 100 at 5°C to 1000 at 120°C (also varying somewhat with  $\sigma^{\alpha\beta}$ ) and, with the above assumption for  $\sigma^{\alpha\beta}$ , they correspond to  $R^l$  increasing from 0.06 to 0.45 over the same temperature range. As already explained the Zener spacing is  $R^l = \frac{1}{2}$  so the observed spacing is three times larger at the higher temperatures and eight times larger at the lower temperatures.

The important kinetic parameters of the incoherent boundary which are required for a detailed test of Cahn's theory are not known, but some internal tests may be made. The above result for  $R^l$  restricts  $\beta^l$  to a small range of values near  $10^{-2}$ , which in turn implies that  $Q^l \cong 0.6$ ; this applies at room temperature where  $R^l$  and  $\beta^l$  should be smaller than the above estimates and  $Q^l$  correspondingly a little larger. However,  $R^l$  depends upon  $\Delta g^l$  and  $\sigma^{\alpha\beta}$  (neither of which is known accurately) so the comparison is inconclusive. The experimental measurements of spacing also cover such a small range of supercooling or supersaturation that  $\Delta g^l$  and  $\beta^l$  do not change sufficiently to produce much variation in  $P^l$  or  $Q^l$ .

According to eqn. (55.24), the growth rate is proportional to  $-\Delta g^l$  and hence to  $-\Delta g^l$  if  $\beta^l$  is small and  $(P^l - R^l)$  nearly constant. Accepting as before that  $\Delta g^l$  varies as  $(x^m - x^\alpha)^2$  for small supersaturations, this gives a square dependence of  $\Gamma$  on the supersaturation. Turnbull's prediction from eqn. (55.20) gives  $\Gamma$  proportional to  $(x^m - x^\alpha)^5$ ; available data seem to support the square rather than the fifth power dependence. Turnbull's expression is only valid if the degree of segregation is not allowed to vary with the conditions of growth. Turnbull and Treafis examined the prediction that  $\Gamma$  varies as  $(y^{\alpha\beta})^{-2}$  and found this to be in reasonable agreement with both their own and Borelius' kinetic data; presumably this is again to be attributed to the comparatively small range of temperatures over which experimental measurements can be made.

Several other discontinuous precipitation reactions have been investigated. Many of the later papers interpret the experimental results in terms of a simple theory by Petermann and Hornbogen (1968) in which the velocity is explicitly expressed in terms of the driving force. Gust (1982) compared different theories by evaluating the product  $Q^{lB} D^B$  and comparing this with the experimental results for diffusion along stationary grain boundaries. For the Turnbull theory, eqn. (55.20) leads to

$$Q^l \delta^B D = [c^m / (c^m - c^\alpha)] \Gamma (y^{\alpha\beta})^2 \quad (74.4)$$

whilst the Petermann and Hornbogen theory gives

$$Q^l \delta^B D = (kT / 8 \Delta g^l) \Gamma (y^{\alpha\beta})^2 \quad (74.5)$$

These two simple equations give quite good agreement with the results for diffusion along static boundaries in the case of Ni–In alloys (Gust *et al.*, 1982) and in general for all eight systems for which data are available (Gust, 1982).

Al–Zn alloys have been much investigated; some complications arise from simultaneous or prior spinodal decomposition (Sun *et al.*, 1988), the occurrence of discontinuous coarsening and the suggestion that the precipitation reaction begins with diffusion-induced grain boundary migration (DIGM). A complication in Al–Zn (Butler *et al.*, 1973; Abdou *et al.*, 1988) and Cu–Be alloys (Alexander *et al.*, 1982) is the occurrence of continuous precipitation as a competing reaction; the amount of discontinuous precipitation is increased by rapid water-quenching and decreased by higher solution temperatures.

Discontinuous coarsening and discontinuous dissolution of the precipitate have both been studied extensively (Abdou *et al.*, 1988; Fournelle *et al.*, 1988). Both reactions occur discontinuously at an interface, and in the coarsening reaction there is usually a further draining of solute from the still supersaturated  $\alpha$  phase. The microstructure undergoes a dramatic change, the interlamellar spacings increasing by a factor of three to eight. The coarsening reaction normally takes place with a much slower velocity than the precipitation reaction (one or two orders of magnitude), but this is presumably to be attributed to a lower driving force as it has the same apparent activation energy as the initial precipitation and thus seems to be controlled also by grain boundary diffusion. This also applies to the discontinuous dissolution process.

#### 75. GP ZONE FORMATION AND OTHER LOW TEMPERATURE CHANGES

It was pointed out in Section 72 that the conductivity, hardness and other physical properties of certain alloys change markedly with time although no precipitates can be detected by normal microscopic methods. These changes, which normally occur only at low temperatures, are associated with the precipitation phenomena usually known as clustering or GP zone formation. They are found in alloys where the atomic size misfit, i.e. the difference in atomic diameters of the components, is small, or the supersaturation is large, so that cluster nucleation is not unduly difficult. Typical examples are the supersaturated solutions of cobalt in copper, and of silver or zinc in aluminium, in all of which the difference in atomic sizes is small. The already discussed “ageing” changes in steels tempered at around room temperature should also logically form part of this section, but they are not usually classed as GP zones for historical reasons. The best known pre-precipitation phenomena are found in solutions of copper in aluminium, where the size disparity is quite large (about 12%) but zones form only at high supersaturations.

Evidence for more than one product in low temperature ageing first came from anomalous changes in properties such as hardness and electrical conductivity. Figure 16.27 shows the change of hardness of an aluminium–silver alloy isothermally annealed at two different temperatures, corresponding (in Köster’s notation) to “cold” and “warm” hardening respectively. Figure 16.28 shows isochronal annealing curves at various temperatures; the softening after the first hardening peak is generally referred

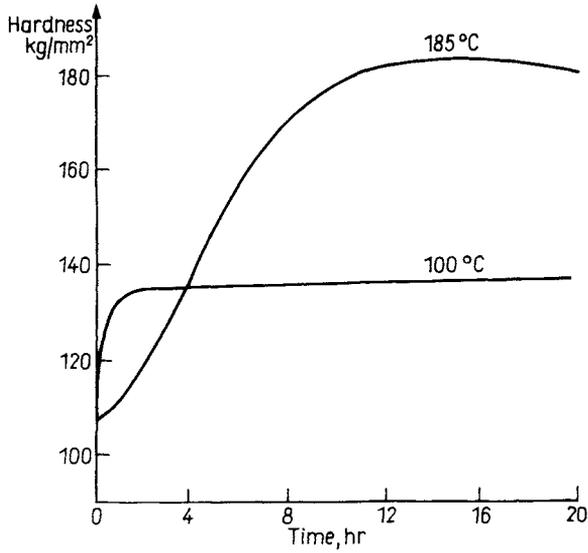


FIG. 16.27. Variation of the hardness of an aluminium-38% silver alloy during warm and cold ageing (after Köster and Braumann, 1952).

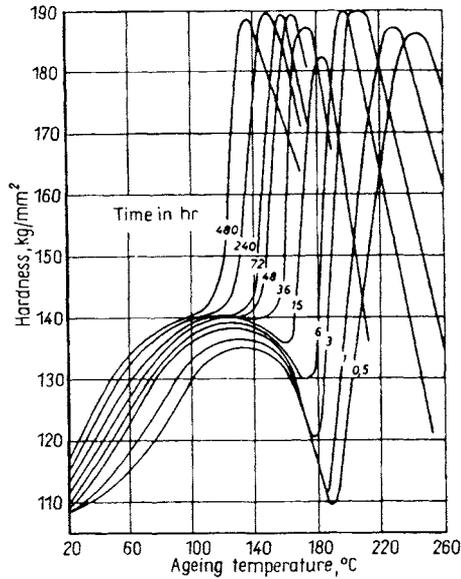


FIG. 16.28. Isochronal hardening curves for an aluminium-38% silver alloy (after Köster and Braumann, 1952).

to as reversion or retrogression and it is utilized in the industrial processing of aluminium alloys.

Consider a supersaturated solution which transforms at relatively high temperatures to give a continuous general precipitate, with each matrix grain retaining its shape and orientation, but not its composition. As the temperature of precipitation is lowered, the precipitate particles become smaller and more numerous; new non-equilibrium precipitates may also form, especially if they can establish coherence with the matrix or first-formed precipitate phase. For a given product phase, however, the transformation structure after long times at various temperatures differs mainly in the distribution of the appropriate particles, and there are only small differences in the amount of solute left in solution. Hence, if a temperature range is reached in which there is no visible precipitation, despite changes in physical and mechanical properties, it is a natural extension to assume that precipitation is again virtually completed. The size of the particles may simply have become too small for detection by optical methods or, because of diffraction broadening, by normal X-ray techniques. After prolonged ageing, the particles often appear, because of coarsening, to reduce the surface free energy.

The above description, if couched in terms of only one precipitate, was known for some time as the simple precipitation theory of age hardening or the sequential theory, and some authors strongly advocated it as the true situation in all alloys. The opposing view, i.e. that changes within the solid solution precede the formation of the first precipitate, and that different precipitate phases may form and disappear again before the equilibrium precipitate is formed, received experimental support from measurements of diffuse X-ray scattering, using single crystals of aluminium copper alloys (Preston, 1938ab; Calvert *et al.*, 1939). For many years, it seemed to be possible to interpret the X-ray evidence using either hypothesis, and some developments of the single precipitation theory claimed to show that the tiny precipitate particles grow anisotropically, first forming needles which thicken into plates and finally into three-dimensional crystals (Geisler and Hill, 1948). It is now known that in most conditions where property changes do not correlate directly with changes in the microstructure, the sequential theory of precipitation is inadequate.

Structural work on the first precipitates or "zones" which form at the lowest temperatures has mainly utilized X-rays and electron microscopy. The ideal GP zone is a region of the matrix with a high solute concentration, but in which all the atoms are located on the sites of the random solid solution. Thus there is no deformation of the lattice around the zone and the matrix and solute atoms must have very nearly equal atomic diameters. Nearly ideal zones are formed in alloys of aluminium with silver or zinc. More usually the size difference of the atoms will produce some distortion, and the atoms in and around the zone will be displaced from the lattice sites of the initial solution. It is characteristic of a cluster, as distinct from a coherent precipitate, that the atoms do not approximate to a new regular arrangement; the displacements may be regarded as large elastic strains of the matrix, and the distortion may vary from the centre of the zone to outside the zone. There is perfect coherence between the zone and the matrix and all the atoms in the zone are perturbed by the constraints of the matrix. It follows that only small zones can be stable, especially if the displacements from the ideal positions are

appreciable. The shapes and sizes of the zone can be defined only approximately as there are no sharp boundaries but the linear dimensions are frequently of the order of 10 nm or less. The zone shape varies in different alloys; in aluminium–silver, for example, it is spherical whilst in aluminium–copper the zones are plate-shaped, as might be expected from the larger elastic strains.

The X-ray data for clustering in aluminium–silver alloys have been analysed in detail. The work of Walker and Guinier showed that the zones have spherical symmetry, at least statistically, and that each zone is surrounded by a solute-impoverished shell. This was subsequently confirmed by transmission electron microscopy and Freise *et al.* (1961) used both techniques in a detailed study of an alloy containing 4.4 at.% silver. After ageing at 125°C, there are about  $2 \times 10^{21}$  zones  $\text{m}^{-3}$  in this alloy, the diameter of each zone varying from about 2 nm to about 6 nm. The spherical clusters contain more than 90% silver and the shells surrounding them are nearly pure aluminium. Figure 16.29 shows a plot of the quantity  $P_{\Lambda_{\text{Ag}} \Lambda_{\text{Ag}}}(r)$  obtained from X-ray data. [ $P_{\Lambda_{\text{Ag}} \Lambda_{\text{Ag}}}(r)$  (see p. 172) is the probability averaged over all silver atoms that the neighbour at distance  $r$  from a silver atom will be another silver atom.] In contrast to some earlier results on Al–Cu alloys, however, Freise *et al.* found that only about 10% of the silver is segregated into the zones, the remainder being in solution. Figure 16.29 is normalized so that it refers only to the part of the alloy which is segregated.

The formation of a spherical cluster requires atomic movements over comparatively small distances, the excess solute atoms in the cluster just balancing those removed from

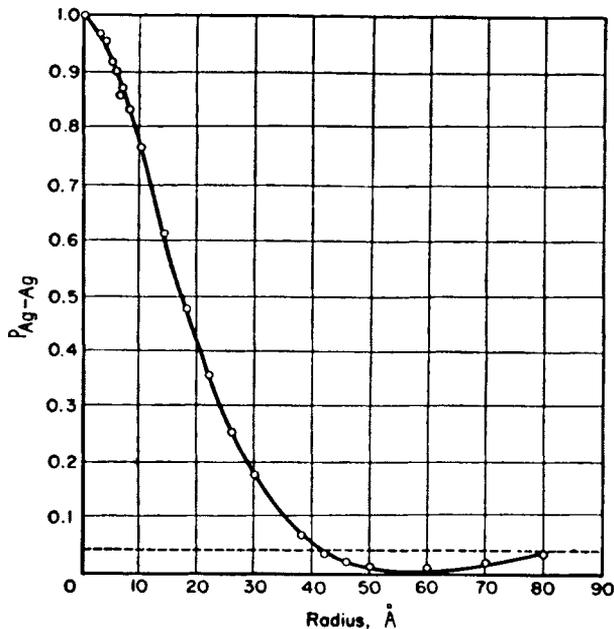


FIG. 16.29. Plot of  $P_{\Lambda_{\text{Ag}} \Lambda_{\text{Ag}}}$  against  $r$  for spherical zones in an aluminium 4.4 at.% silver alloy (after Freise *et al.*, 1961).

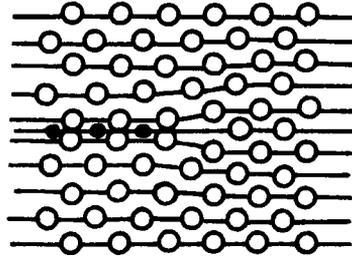


FIG. 16.30. Model of GP zone in aluminium–copper alloys (after Gerold, 1954).

the surrounding shell. A qualitative picture of the process is that an unstable cluster embryo grows by addition of atoms from its immediate neighbourhood, the uphill diffusion being possible because of the lower free energy of the cluster. Normal downhill diffusion in the matrix should tend to remove the depleted region, but this is a slow process in the diluted solution. At a certain diameter of the depleted zone, diffusion will effectively cease, the probability of a solute atom being able to cross the depleted zone in a reasonable time then being negligible. As atoms are more mobile at higher temperatures, the zones will become larger and the depleted regions less well-defined until the latter have effectively disappeared. At still higher temperatures, the zones become unstable against the disordering effects of thermal vibrations and they dissolve again (reversion). This reversion temperature is effectively the solvus line for the GP zones. Only aggregates of solute atoms which have a distinct crystal structure have sufficiently low free energies to remain stable above the GP solvus.

In aluminium–copper alloys, there is no evidence of a well-defined impoverished region adjacent to the platelets. The greater rate of diffusion which this implies may be due to the distortion around the zone. Many models incorporating this distortion have been proposed. A schematic model due to Gerold (1954) is shown in Fig. 16.30. There is a single plane of copper atoms succeeded by planes containing mainly aluminium atoms and the distortion extends through many planes of the matrix. This gives good agreement with the diffraction results although it is difficult to deduce the exact dimensions of the zone. Guinier (1959) pointed out that other models with changing ratios of copper to aluminium are also able to account for the experimental results.

The size of a GP zone is greater than the resolving power of a modern electron microscope, and the application of the thin film technique to study zones was pioneered by Castaing (1956), although later work has shown that some of his photographs were wrongly interpreted. A thin film transmission electron micrograph of the first detectable zones in aluminium–copper alloys is shown in Fig. 16.31 taken from the work of Nicholson and Nutting (1958). The discs can be seen in one orientation only and appear as needles 0.4–0.6 nm thick and 8 nm long.

A somewhat different picture emerges from a field ion microscope study by Abe *et al.* (1982) who found several layers in the three-dimensional structure (Fig. 16.32) which they



FIG. 16.31. Transmission electron micrograph of GP I zones in an aluminium copper alloy (original magnification  $\times 800\,000$ ; courtesy of Nicholson and Nutting, 1958).

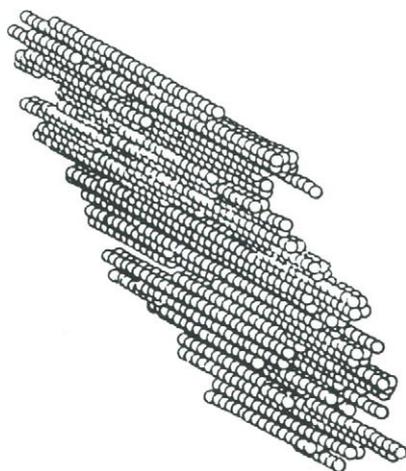


FIG. 16.32. Model of GP I zone obtained from field ion microscopy (after Abe *et al.*, 1982).

describe as copper-rich plates 4 nm in average diameter with inhomogeneous thickness and complex shapes containing ledges and holes.

The plate-shaped zones shown in Fig. 16.31 are known as GP1 and are responsible for the change of properties produced by ageing at low temperatures, as shown for example by the cold-hardening curve of Fig. 16.27. On ageing for long periods at temperatures in the region of 100°C, or for shorter times at higher temperatures, the hardness curve rises to a second maximum, followed eventually by a decrease. The X-ray patterns show that, during this period, there are changes in the matrix resulting in the formation of a new structure which is variously described in the literature as  $\theta''$  or GP2. The first term is due to Guinier (1942, 1950) who used it to describe an intermediate ordered structure forming after the GP1 zones and prior to the coherent non-equilibrium phase  $\theta'$ . According to Guinier, the  $\theta''$  regions may consist of layers parallel to the matrix {100} planes, the central plane being pure copper, followed by adjacent planes 0.19 nm away composed of mixed copper and aluminium atoms, and then planes another 0.2 nm away composed of essentially pure aluminium. This gives formally a tetragonal structure with  $a = 0.404$  nm and  $c = 0.79$  nm, but not all of the observed diffraction effects can be explained in this way, and it may be necessary to double the  $c$  parameter to obtain the correct unit cell. This parameter varies with the details of the heat treatment, being 0.8 nm when first formed but decreasing to 0.76 nm as the zones become larger (Silcock *et al.*, 1953–4). The formation of  $\theta''$  is detected in X-ray images by the splitting of the streaks from GP1 into discrete spots. Nicholson and Nutting's transmission electron microscopy images of the GP2 phase and the  $\theta'$  phase are shown in Figs. 16.33 and 16.34 respectively. In Fig. 16.34, the foil surface is near (100) and the plates appear as needles of 2 nm thickness and 30 nm length. Comparison of electron diffraction patterns from such an alloy with those containing GP1 or  $\theta'$  particles indicates that GP2 is much better regarded as a genuine precipitate, so that the name  $\theta'$  seems more appropriate. However, a later weak beam electron microscopic study by Yoshida (1982) suggests a progressive build-up from GP1 zones, which are modelled much as in Fig. 16.32, via GP2 zones, which have two planes of copper atoms separated by a plane of aluminium atoms, and "GP3" zones, with three planes of copper atoms, to the structure of  $\theta'$ , which is the first clearly recognizable precipitate phase.

An aluminium alloy hardened at low temperatures by precipitation of GP zones or  $\theta''$  begins to soften again when the  $\theta'$  precipitate is visible. On annealing above the reversion temperature,  $\theta'$  is the first structure to form, and leads to hardening followed by eventual softening. The  $\theta'$  structure is also not an equilibrium phase but a coherent intermediate precipitate of tetragonal structure, and has a fixed orientation relation with the matrix. The particles probably have a semi-coherent interface of the van der Merwe kind with the matrix (see p. 363) and the dark ring around each particle in Fig. 16.34 may indicate the presence of interface dislocations. Interpretations of this kind must be made with caution, however, because of the complex diffraction contrast from precipitates; the contrast inside the particles of Fig. 16.34, for example, is probably associated with surface steps.

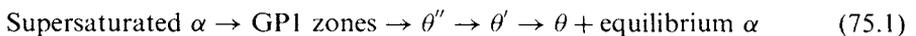
The composition of  $\theta'$  is probably close to that of the final equilibrium precipitate  $\theta$  or  $\text{CuAl}_2$  although, according to Silcock and Heal (1956), a more probable formula for  $\theta'$  is  $\text{Cu}_2\text{Al}_{3.6}$ . The  $\theta'$  structure is the first precipitate which may be recognized under the optical



FIG. 16.33. Transmission electron micrograph of  $\theta'$  (or GP 2) in an aluminium-copper alloy (original magnification  $\times 200\,000$ ; courtesy of Nicholson and Nutting, 1958).

microscope and, at high temperatures, at least, it forms directly from the matrix, as does the precipitate in other continuous reactions. The equilibrium precipitate  $\theta$  is also tetragonal; it is not coherent with the matrix.

The structural evidence on the low temperature precipitation processes in aluminium-copper alloys thus leads to the conclusion that there can be at least four distinct stages:



It is not known in all cases whether these structures form separately or are nucleated independently. Guinier (1942) showed that  $\theta$  may form either by a transformation of existing  $\theta'$  or directly from the matrix, depending on the degree of supersaturation. Silcock *et al.* (1953-4) found that under suitable conditions either GP zones,  $\theta''$  or  $\theta'$  may be the first detected structure, suggesting that it is at least possible for them to be independently nucleated. Because of the difference in scale (Figs. 16.31 and 16.33) the transition from GP1 zones to  $\theta''$  must require that most of the former redissolve, so that it is not very meaningful to ask whether the latter form at the segregates of the first stage.

The structural changes occurring in this alloy system have been described at some length in order to illustrate the complexity of low temperature precipitation phenomena. Although understanding of the various stages is still incomplete, good progress has been made in relating the structures to the property changes. Figure 16.35 taken from Silcock

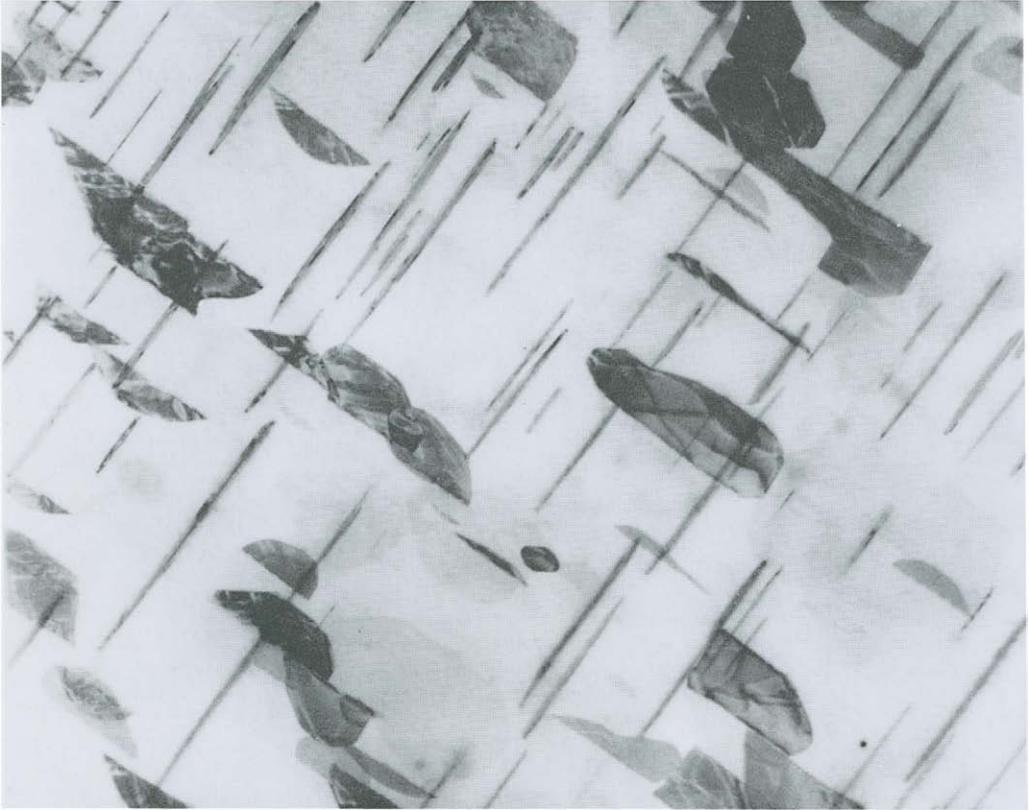


FIG. 16.34. Transmission electron micrograph of  $\theta'$  precipitate in an aluminium-copper alloy (original magnification  $\times 21\,000$ ; courtesy of Nicholson *et al.*, 1958-9).

*et al.* shows the correlation between zone diameter and hardness in three aluminium-copper alloys aged at  $110^{\circ}\text{C}$ .

For other alloys showing low temperature precipitation and age-hardening effects, knowledge of the structural changes is less complete. In aluminium-silver alloys, there is a coherent metastable product which precedes the formation of the equilibrium precipitate ( $\gamma$ ). In the work of Guinier and his collaborators, the first stage was found to be the formation of the spherical silver-rich clusters already described. Walker and Guinier (1953) found such clusters to be present immediately after quenching, so there is reason to believe that they are formed during the quench, or are present even at the pre-quenching temperature (Walker *et al.*, 1952). Support for the conclusion that some clustering or pre-precipitation process occurs in these alloys at a rate too rapid to be suppressed by quenching also comes from the observation (Borelius, 1954; Köster and Knodler, 1955) that the electrical resistance immediately after quenching is much smaller than the value extrapolated from high temperatures. However, Turnbull and Treafis (1957) succeeded in

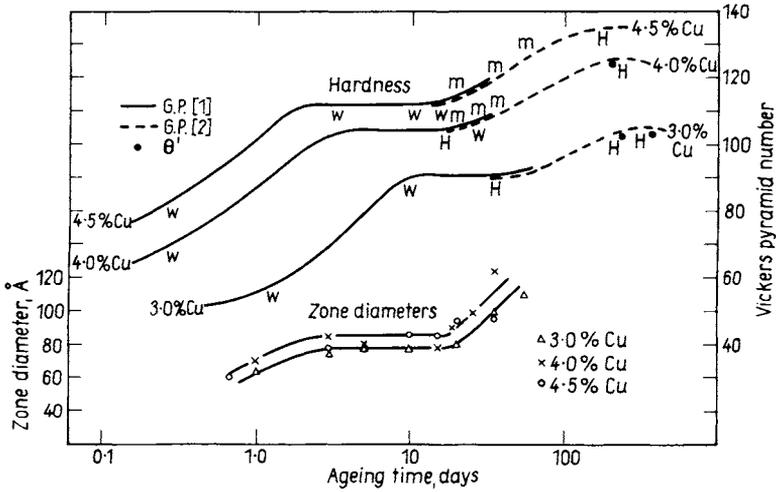
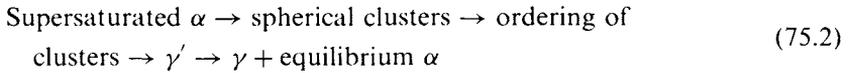


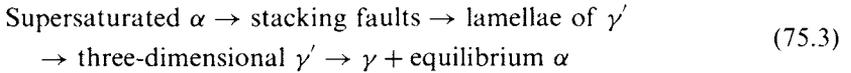
FIG. 16.35. Structures and zone diameters in three aluminium-copper alloys aged at 110 C in relation to curves of hardness versus ageing time (after Silcock *et al.*, 1953 4).

quenching alloys containing 2.5% and 8.5% silver sufficiently rapidly that their resistances did lie on the extrapolated high temperature curves.

The French work leads to the conclusion that the structural changes in aluminium-silver alloys may be described by a sequence similar to that found in aluminium-copper alloys:



Investigations by German authors show that this sequence does not represent all of the changes which may take place (Glocker *et al.*, 1952; Köster and Braumann, 1952; Ziegler, 1952). These authors show that, during ageing, stacking faults may develop in the matrix, followed by diffusion of solute atoms to the faults which thus become coherent platelets of  $\gamma'$ . There is thus an alternative sequence:



Some of the differences between sequences (75.2) and (75.3) are attributable to the compositions of the alloys; the French and American workers mainly used alloys with less than 20 wt.% silver whereas the German workers used a 38% silver alloy. Nevertheless, it seems probable that both sets of reactions occur in alloys of all compositions and this view has been confirmed by the electron microscopy work of Nicholson and Nutting (1961). This investigation showed that, for a 16 wt.% silver (4.4 at.%) alloy, the basic precipitation sequence (75.2) is correct, but complications are caused by the tendency of the silver-rich phase to order, and by the lowering of the stacking fault energy in the

aluminium–silver solid solution. The quenching process introduces helical dislocations (see p. 252) which absorb silver to produce long, narrow stacking faults on {111} planes. These faults serve as nuclei for the precipitates which are thus formed independently of the GP zones. The GP zones themselves grow continuously during ageing unless a plate forms near them; ultimately they redissolve in favour of the  $\gamma$ . This is in contrast to the behaviour of GP1 zones in aluminium–copper alloys which grow to a certain size and then remain unchanged for some time whilst  $\theta'$  is precipitated. Above the reversion temperature, many of the clusters dissolve in the usual way whilst others grow to larger sizes.

Structural investigations of a detailed nature have also been made on aluminium–magnesium–silicon, aluminium–zinc, aluminium–magnesium and copper–beryllium alloys, amongst others. The complex and varied results obtained will not be described, however, as they reveal no new mechanisms. The formation of GP zones in copper–aluminium alloys releases up to  $17 \text{ kJ g atom}^{-1}$ , and this is of the same order as the energy released in the high temperature precipitation of the equilibrium phase. During isothermal zone formation, the rate of release of enthalpy is a maximum initially and continually decreases, so that the curves of energy versus time have the same form as the cold-hardening curve (Fig. 16.27). This behaviour is in marked contrast to the sigmoidal reaction curves characteristic of most nucleation and growth transformations and exhibited by precipitation at higher temperatures in age-hardening alloys.

Turnbull (1956) suggested that a possible explanation of the kinetic law is a critical nucleus size for the zones of less than one atom. Isothermal reaction will then consist entirely of the growth of some zones at the expense of the others, the smaller zones dissolving to allow the larger ones to grow. Such coarsening under the influence of surface free energy is a feature of most precipitation reactions and is discussed in Section 76.

Jagodzinski and Laves (1949) recognized that the formation of GP zones in aluminium–copper alloys at room temperature takes place at rates many times (10–17) faster than would be predicted from the extrapolated diffusion coefficient. Seitz (1952) suggested that diffusion rates might be enhanced by excess vacancy concentrations introduced by the quench, or formed by plastic deformation. Another possibility (Turnbull, 1954) is that dislocation lines might enable the solute atoms to drain to the clusters. Stationary dislocations would not be effective but, if dislocations glide during the process, it is possible to account for the results in a manner analogous to that used to describe discontinuous precipitation. Fisher and Hollomon (1955) suggested that the change in free energy due to the formation of GP zones will produce a stress which could cause dislocations to glide, but no detailed mechanism has been suggested.

De Sorbo *et al.* (1958) designed experiments to give more information on the mechanism of the rapid low temperature formation of clusters. Using a 1.9 at.% copper alloy, they found that appreciable release of energy (up to  $250 \text{ J g atom}^{-1}$  of alloy or  $14 \text{ kJ g atom}^{-1}$  of copper) and increases in electrical resistivity (up to 5%) can occur isothermally at temperatures as low as  $-45^\circ\text{C}$ . The changes were attributed to GP zone formation, or to

the formation of clusters preceding the zones detectable by diffraction techniques. The change in energy and resistance could be satisfactorily fitted to an equation

$$dU/dt = C_{34}/(t + C_{35}) \quad (75.4)$$

The effect of varying the cooling rate was to produce more rapid transformation after faster quenches. When the quench was interrupted holding for a few seconds at 200°C, the transformation rate was decreased by a factor of 10–100 although the holding itself produced no change in the electrical resistivity.

Analysis of the electrical resistivity curves using an equation of the form of (75.4) showed a temperature variation of the type

$$C_{34}/C_{35} = C \exp(-Q/kT) \quad (75.5)$$

with an activation energy  $Q$  of just under 5 kJ g atom<sup>-1</sup>, or about 0.5 eV/atom. This energy is about equal to that required to anneal out point defects in quenched pure aluminium (DeSorbo and Turnbull, 1959) and is close to the corresponding energy for clustering in an aluminium–silver alloy (see below). The results may be interpreted on the basis of the vacancy theory mentioned above, which gives an obvious interpretation of the activation energy and of the results of the interrupted quench. The vacancy hypothesis is also consistent with some other observations; for example, light cold working increases the rate of precipitation in alloys which have not been cooled very rapidly, but the rate never exceeds that of the most rapidly cooled specimens. At very low temperatures, the rate is independent of thermal cycling. The rate of movement of copper atoms in these experiments is at least 10<sup>13</sup> times greater than that given by the extrapolated high temperature lattice diffusion coefficient. Rather artificial assumptions have to be made in order to fit the results to the dislocation hypothesis.

As described in Chapter 7, electron microscopy studies on quenched metals have shown directly that dislocation loops or stacking fault tetrahedra are formed by the collapse of vacancy discs. Slightly different results have been obtained for quenched alloys. In aluminium–copper alloys, for example, loops are found only up to about 2% copper, and at higher copper contents large numbers of helical dislocations are observed. These must result from the condensation of point defects on screw dislocations, presumably because the solute atoms interact strongly with the vacancies and prevent their forming dislocation loops. In a study of the structures in a quenched 4.5% copper alloy, Thomas and Whelan (1959) calculated that the number of vacancies which reach the dislocation helices during the quench is much smaller than the number of excess vacancies, so that an appreciable fraction is retained in solution.

A difficulty with the vacancy theory is that the normal lifetime of quenched-in point defects is very short in comparison with the ageing periods of hours, days or even weeks. Thomas and Whelan's data supports the possible explanation, first suggested independently by Panseri *et al.* (1958) and by Hart (1958), that the vacancies are retained in solution for long periods because of the strong attractive force between vacancies and solute atoms. Hart suggested that two solute atoms interact more strongly with a vacancy than one, so that this little cluster can move through the lattice at an appreciable rate. As it

does so, it gradually acquires more solute atoms and vacancies, until the aggregates thus formed have mopped up all the excess solute. The vacancies thus do not escape to sinks but are available to assist movement over long periods. As some vacancies clearly condense on screw dislocations to give helices, they must presumably take solute atoms with them. The solute atoms should interact fairly strongly with the edge components of these helices, which thus become favoured sites for nucleation of precipitate. Thomas and Nutting (1959) have presented evidence to show that the spacing and distribution of  $\theta''$  particles is consistent with their having been nucleated on successive turns of a dislocation helix.

The kinetics of low temperature cluster formation were studied in aluminium–silver alloys by Turnbull and Treafits (1957). After very rapid quenching, alloys containing 2.5% and 8.5% silver were found to have the electrical resistance expected from the high temperature results, in contrast to the results of earlier workers (see p. 780). The resistance of these alloys was found to decrease with time on isothermal treatment at temperatures as low as  $-30^\circ\text{C}$  and  $-60^\circ\text{C}$  for the two compositions respectively. Figure 16.36 shows a set of resistance versus time curves for the 2.5 at.% silver alloy. The rate of change of resistance with time was fitted to an equation of the form

$$d\Delta P/dt = C_{36}/(C_{37} + t)^{2/3} \quad (75.6)$$

The isotherms for each composition differ only by a constant factor  $f$  multiplying the time scale;  $\ln f$  is linear with  $1/T$  as shown in Fig. 16.37 and the activation energies deduced from this relation (50 and 47 kJ g atom $^{-1}$ ) are both comparable with the corresponding

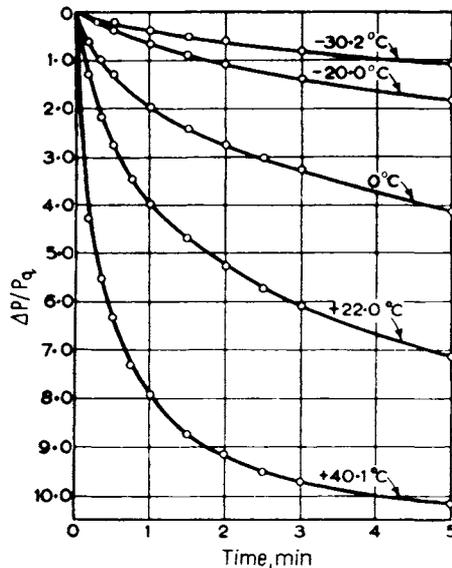


FIG. 16.36. Isothermal curve of electrical resistivity versus time for an Al-2.5 at.% Ag alloy (after Turnbull and Treafits, 1957).  $\Delta P = P_t - P_0$ , where  $P_t$  and  $P_0$  are the resistivities at times  $t$  and zero respectively and  $P_q$  is the resistivity at 273 K and  $t = 0$ .

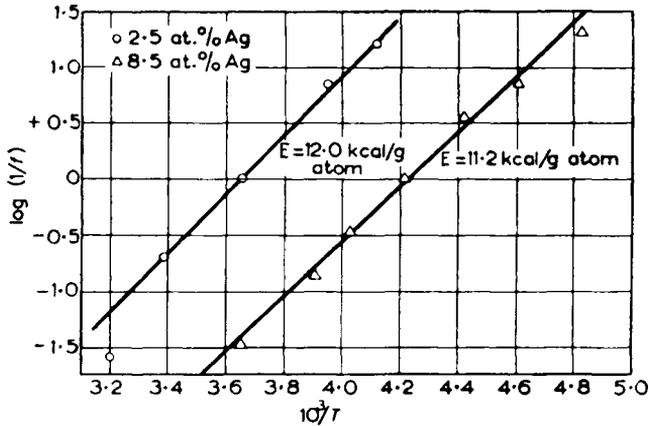


FIG. 16.37. The temperature dependence of the timescale factor  $f$  for two aluminium-silver alloys (after Turnbull and Treafitis, 1957).

activation energy for zone formation in aluminium-copper alloys. The results for these alloys are thus consistent with the hypothesis that the very rapid transformation rate is due to excess vacancy concentrations. The rates at 22 K are too rapid by a factor of  $10^{15}$  for normal lattice diffusion.

#### 76. COARSENING AND ELASTIC INTERACTION BETWEEN PRECIPITATES

After a precipitate reaction has ended, it is commonly found that the microstructure continues to change because of the interfacial free energy and the strain energy which are contained in the structure. The influence of the surface free energy is to coarsen the structure by increasing the number of larger-sized particles and decreasing the number of smaller particles thus, at fixed partition of solute between the two phases, decreasing the total interfacial area and hence the free energy. This process is known as Ostwald ripening, and was first treated theoretically by Lifshitz and Slyozov (1961) and independently by Wagner (1961). Their rather complex theory has been reviewed several times (Greenwood, 1968; Ardell, 1988; Doi, 1996) and only an outline treatment will be reproduced here. Transfer of solute from the smaller particles to the larger occurs by volume diffusion in the matrix, the composition gradient being provided by the dependence of the concentration in the matrix immediately adjacent to a particle on the radius of curvature of that particle, i.e. by the Gibbs-Thomson effect. The LSW theory also assumes that the particles have a spherical shape so that the concentration adjacent to a particle of radius  $r$  is

$$c(r) = c(\infty)[1 + 2\Gamma/r] \quad (76.1)$$

[see eqn. (22.34)]. Under the assumption that the mean separation of the particles exceeds the average particle size, the rate of change of the radius because of diffusion

may now be written

$$dr/dt = (2D/\Gamma r)[(1/\bar{r}) - (1/r)] \quad (76.2)$$

where  $\bar{r}$  is the radius of spheres which are neither growing nor shrinking. In the limit of long ageing times

$$\bar{r}^3 - \bar{r}_0^3 = 8\Gamma Dt/9 \quad (76.3)$$

where  $\bar{r}_0$  is the mean radius in the initial distribution. The particle size distribution is finally

$$\begin{aligned} f(r, t) &= [A/(1 + t/t_d)^{4/3}] \rho^2 [3/(3 + \rho)]^{7/3} [3/(3 - 2\rho)]^{11/3} \exp[-2\rho/(3 - 2\rho)] \\ &= A\rho^2 h(\rho)/(1 + t/t_d)^{4/3} \end{aligned} \quad (76.4)$$

where  $A$  is a constant and  $t_d$  is given by

$$t_d = 9kT\bar{r}_0^3/8Dv^2\sigma c(\infty)$$

for  $\rho = r/\bar{r} < 3/2$ . For  $\rho > 3/2$ ,  $f(r, t) = 0$  in the LSW theory.

Equations (76.3) and (76.4) constitute two of the main results of the LSW theory, namely that the mean radius  $r$  increases as  $t^{1/3}$  and that the distribution at any time cuts off at  $1.5\bar{r}$ . In general, it is found experimentally that the first prediction is well obeyed, although there are exceptions, but the distribution of sizes is much broader than is given by eqn. (76.4). There is general agreement that the discrepancy arises, at least in part, from the assumption of the theory that the volume fraction of precipitate is negligible, so that the particles are well separated in a liquid-like matrix. Several attempts have been made to modify the theory to allow for a finite volume fraction. The theory of Ardell (1972) allows for the interactions of overlapping diffusion fields and predicts the growth law as

$$\bar{r}^3 - \bar{r}_0^3 = K(\phi)t$$

where  $K(\phi)$  is a rate constant and is a function of the particle volume fraction  $\phi$ . The theory preserves the  $t^{1/3}$  growth law, as does an alternative description by Davies *et al.* (1980) which focused attention on another effect of a finite volume fraction, namely a greater tendency for growing particles to coalesce; this is called the encounter theory. Other "ad hoc" theories have been advanced by Asimov (1963), by Sarian and Weart (1969) and by Aubauer (1978) and the theory has subsequently been surveyed by Ardell (1988) and by Doi (1996).

More rigorous theories attempt to calculate  $\phi$  in a statistically averaged sense, taking into account that the local environment of a particle of any size will vary over the ensemble. Thus a particle of size  $\bar{r}$  may locally be either growing or shrinking, but according to eqn. (76.2) it is doing neither. General theories dealing with the problem have been developed by Brailsford and Wynblatt (1979), Voorhees and Glicksman (1984ab), Marquese and Ross (1984), Tokuyama and Kawasaki (1984) and Enomoto *et al.* (1986); Kawasaki and Tokuyama (1987), Ardell (1988) and Doi (1996) have discussed and compared the various theories.

Computer simulation has been used by many workers in an attempt to sidestep the limitations of the analytical approach. In a number of cases, it has been shown that the  $t^{1/3}$  law is valid and is independent of volume fraction. The simulations encounter the great difficulty that the number of particles that can be handled is very small [four to nine in the initial calculations of Weins and Cahn (1973); 320 in the work of Voorhees and Glicksman (1984)]. Even if the number is adequate at the beginning of the simulation, it diminishes during the simulation so that statistically significant results are difficult to obtain. Enomoto *et al.* (1986) used a trick method of calculation and were able to obtain results from a more realistic starting point of  $10^5$  particles. They concluded that, for volume fractions up to 0.1, their results on the standard deviation of the distribution and on its "skewness" agreed well with the predictions of Tokuyama and Kawasaki (1984) but not with those of Marquese and Ross (1984).

A further difficulty in comparing theory with experiment is that, in many of the experimental results, the particles are far removed from spherical shapes, and in some cases the centres of gravity of the particles are displaced during the coarsening. The coarsening not infrequently overlaps the precipitation processes so that the volume fraction of the precipitate is changing and it is difficult to separate the two effects. The first studies of coarsening (Livingston, 1959) were on cobalt particles dispersed in a copper matrix and the variation of mean particle radius with time, determined by magnetic measurements, obeyed the  $t^{1/3}$  law. By far the greatest number of measurements have been made on  $\gamma'$  precipitates in Ni-Al alloys. This is a coherent precipitate which might be expected to lead to difficulties but in fact the  $t^{1/3}$  law is remarkably well obeyed. Figure 16.38 shows some results of Ardell and Nicholson (1966) for an alloy of

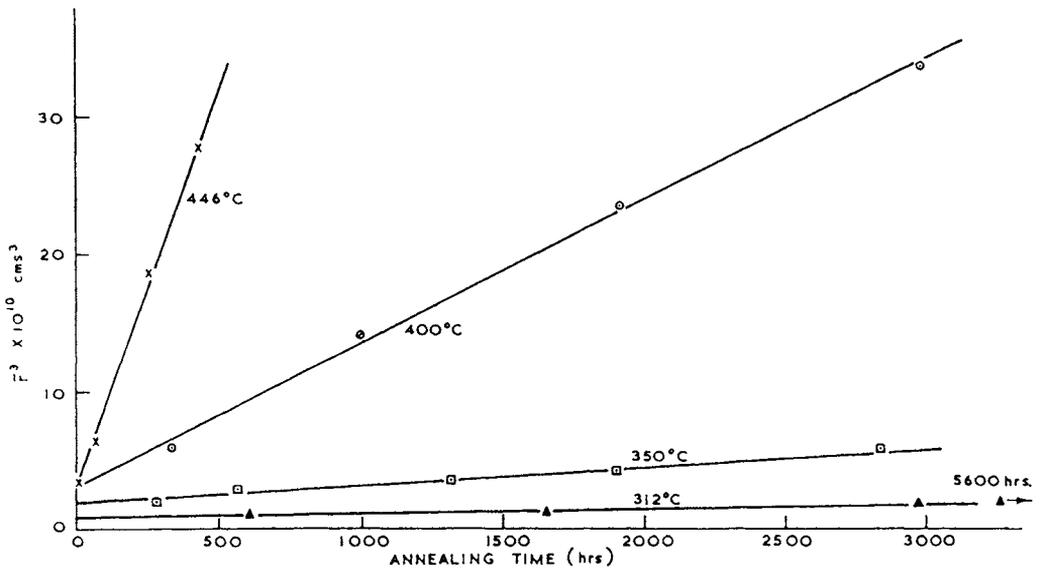


FIG. 16.38. Plot of half the mean particle edge length against  $(\text{time})^{1/3}$  for an alloy of Cr 9.7 wt.% Ni at three different temperatures (after Ardell and Nicholson, 1966).

Cr-9.7 wt.% Ni at three different temperatures. The results of the distribution are not so convincing; the cut-off occurs at twice the mean size rather than 1.5 times, and the distribution is generally broader than predicted. This difficulty has persisted in other measurements, so much so that Ardell concluded his 1988 review entitled "Precipitate coarsening...chronic disagreement with experiment" with the statement "... no one theory is capable of a complete and consistent prediction of the coarsening behaviour of precipitates in any alloy system that has been studied to date." This is in marked contrast to his 1968 paper, which he boldly titled "Experimental confirmation of the Lifshitz-Wagner Theory."

The specific behaviour of coherent particles during coarsening may be expected to be influenced by the elastic interaction as well as by the interfacial free energy. The shape of such a precipitate may change as it coarsens, both because of the influence of its own self energy and also because of its possible elastic interaction with other neighbouring particles. Many of the results for Ni-based alloys do not show coherency effects because the mismatch is small; for example, Doi *et al.* (1996) worked with Ni-18.2 at.% Cr-6.2 at.% Al and Ni-7 at.% Si-6.0 at.% Al alloys for which the misfit strains were about 0.008% and 0.10% respectively. Surface energy rather than strain energy was thus dominant and plots of  $\log r$  versus  $\log t$  gave straight lines with slopes 0.33 and 0.32, consistent with the LSW prediction of 0.33 [see Fig. 16.39(a)]. The standard deviation of the size distribution is shown in Fig. 16.39(b); the measured standard deviations of 0.25

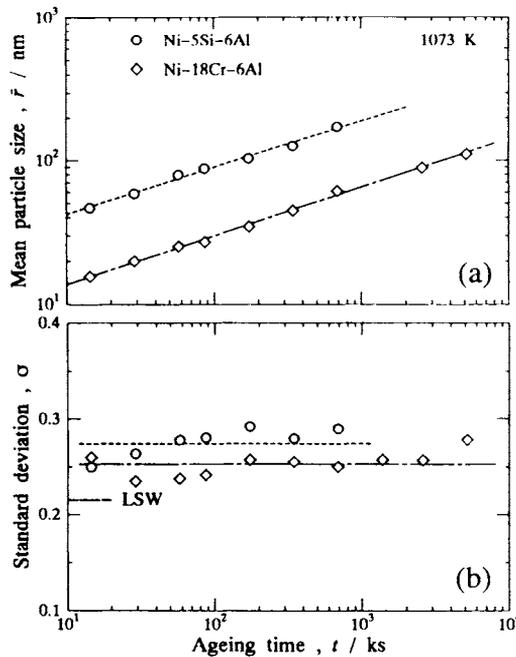


FIG. 16.39. Coarsening kinetics (a) and the standard deviation of the size distribution (b) for two alloys: Ni-Al-Si and Ni-Al-Cr (after Doi, 1996).

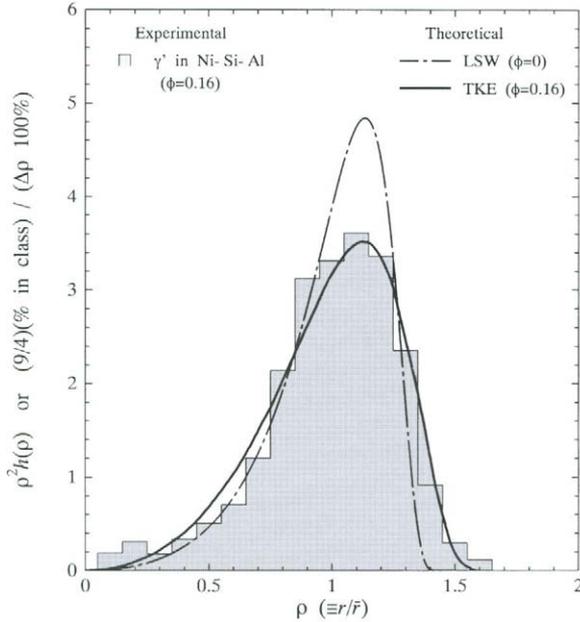


FIG. 16.40. Experimental results for the size distribution curve  $F(r)$  versus  $r$  compared with the predictions of the LSW theory and the theory of Tokuyama and co-workers (after Doi, 1996).

and 0.27 are larger than the LSW prediction of 0.215. However, quite good agreement was observed between the measured distribution and that predicted by the theory developed by Tokuyama (see Fig. 16.40). No difficulties arose with particle shape because the individual particles were spherical.

A relatively large lattice mismatch of  $\sim 1.3\%$  is present in Ni-36.1 at.% Cu-9.8 at.% Si and in Ni-47.4 at.% Cu-5.0 at.% Si. The particle shape is cubic in these alloys and the quantity measured instead of the radius is the half edge of the cube. The coarsening kinetics were found by Miyazaki *et al.* (1988) to give linear  $\bar{r}$  versus  $\log t$  plots but with slopes of 0.28 and 0.17 for the two alloys. The slope decreases with increasing lattice mismatch and also with increasing volume fraction of precipitate. Long-term ageing markedly slows down the coarsening of  $\gamma'$  particles in Ni-Cu-Si alloys. Another difference from the behaviour of alloys with small misfit is that as ageing proceeds the standard deviation gradually decreases; i.e. particles with large or small sizes relative to the mean gradually disappear and the distribution becomes more peaked.

Similar results have been obtained for  $D0_3$  type precipitates, in a b.c.c. matrix. The lattice mismatch is small for Fe-Al-Cr and Fe-Al-Ge alloys and large for Fe-Si-V alloys. The precipitates form as spheres in Fe-Al-Cr and Fe-Al-Ge alloys and as cubes in Al-Si-V. The Fe-Al-Ge alloys behave in accordance with the LSW predictions, with a slope of 0.334, but the coarsening of Fe-Si-V alloys virtually ceases after a certain stage.

In alloys which have a relatively large lattice mismatch, the distribution of precipitate particles is not homogeneous; instead the particles are aligned along  $\langle 100 \rangle$  directions (see Fig. 16.41). Similar microstructures have been observed in the Al-Si-V alloys. The alignment is a result of the elastic interaction between particles, the  $\langle 100 \rangle$  directions

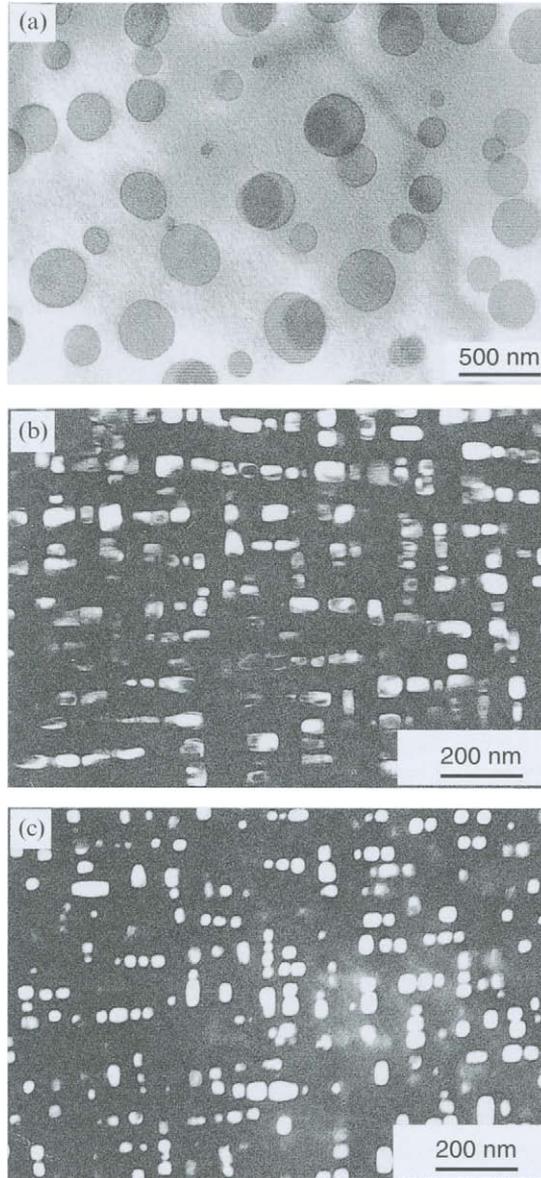


FIG. 16.41. Alignment of precipitates in (a) Fe-13.5 at.% Al-4.4 at.% Ge; (b) Fe-8 at.% Si-8 at.% V; and (c) Fe-5 at.% Si-10 at.% V aged at 923 K (after Doi, 1996).

being the softest elastic directions for most metals. The alignment increases during coarsening.

Another effect of a strong elastic constraint on a coherent precipitate is the splitting of the precipitate particle, either into two parts or into eight parts. Figures 16.42 and 16.43 are examples of doublets and octets respectively observed by transmission electron microscopy. The splitting was first found in  $\gamma'$  precipitates in a Ni-12 at.% Al alloy by Miyazaki *et al.* (1982). Splitting is most readily observed when there is a very low density of precipitates, obtained for example by slow cooling to just below the solvus line and then

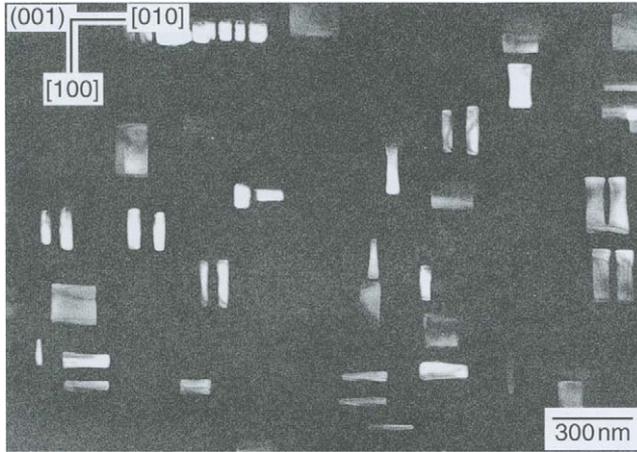


FIG. 16.42. Transmission electron micrograph of  $D0_{22}$  precipitates in an Fe-8 at.% Si-8 at.% V alloy aged at 993 K for 18 ks. The formation of each doublet is the result of the splitting of a single cube (after Doi, 1996).

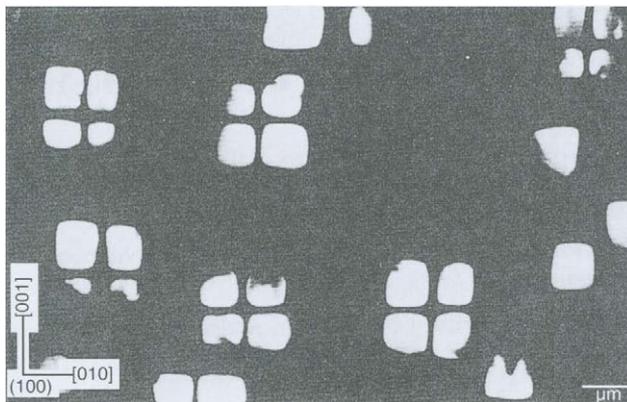


FIG. 16.43. Transmission electron micrograph of  $\eta'$  particles in a Ni-12 at.% Si alloy aged at 1103 K for 72 ks. An octet of cubes results from the splitting of a single cube (after Doi, 1996).

annealing at this temperature for a long period. The energetic situation can be summarized by calculating the interaction energy between two adjacent particles; as the self energies of the two particles equals that of the unsplit original particle, the difference in elastic energies of the two situations depends only on this interaction energy. At sufficient separation, the elastic interaction energy is negative for particles aligned along  $\langle 100 \rangle$  so that the energy of the doublet is given by

$$E^{\text{DBL}} = E^{\text{SNG}} + E^{\text{INT}} + \Delta\sigma$$

Note that because  $E^{\text{INT}}$  is negative for particles aligned along soft directions (usually  $\langle 100 \rangle$ ), the total energy may be reduced by splitting as shown in Fig. 16.42. Thus at a certain size of particle it becomes energetically favourable to split into a doublet.

The corresponding calculation for the split into an octet of smaller particles cannot be easily made because of the difficulty in calculating the elastic energy for the octet. However, an approximation can be used in which a series of pairwise interaction energies with nearest and second and third nearest neighbours is used, so that the interaction energy is

$$E^{\text{INT}} = 12E^1 + 12E^2 + 4E^3$$

This gives the same kind of result, in that the total energy of the octet becomes smaller than the unsplit energy above a certain particle size.

Doi and Miyazaki (1992) introduced a parameter  $\Delta' = \varepsilon/\sigma$ , the ratio of the misfit strain to the surface free energy, to represent the relative contributions to the total energy. According to Doi (1996), for  $\Delta < 0.2$  the surface energy is dominant and no splitting is observed. This is the case for Ni-7 at.% Al, Ni-7 at.% Si, Ni-20 at.% Cr-10 at.% Al and Inconel 50. For  $0.2 < W' < 0.4$ , splitting into octets should occur; e.g. in Ni-10 at.% Al-4 at.% S, Ni-12 at.% Si, Ni-8 at.% Al-5 at.% Ti and Nimonic 115. Finally, for  $\Delta' > 0.4$  the split into two particles will occur; this applies to Ni-12 at.% Al, Ni-11 at.% Ti, Ni-40 at.% Cu-6 at.% Si and Ni-18 at.% Cr-5 at.% Si. The question of the elastic effect and the morphology of the product has also been considered by Khachaturyan *et al.* (1988). Their results, expressed in terms of a parameter  $r_0$  which is defined as the ratio of the surface free energy to the unsplit energy, are as follows:

- $2a > 7.7r_0$ : transition from single sphere to single cube
- $2a > 27r_0$ : transition from single cube to doublet of plate
- $2a > 50r_0$ : transition from single cube to octet of cubes
- $2a > 82r_0$ : transition from doublet of plates to octet of cubes

It should be emphasized again that these transitions will not occur if there is an appreciable density of precipitate so that interactions between existing particles, leading for example to alignment, have already taken place. Moreover the term "splitting" has no mechanistic significance; the effect must depend upon diffusion in the local strain field.

## REFERENCES

- AARONSON, H. I. (1962) *Deformation of Austenite by Diffusional Processes*, Interscience, New York; (1984) *Phase Transformations in Ferrous Alloys*, eds. A. R. Marder and J. L. Golstein, p. 387, Metals Society of AIME, Warrendale, PA.
- AARONSON, H. I. and DOMIAN, K. A. (1966) *Trans. Metall. Soc. AIME* **236**, 76.
- ABDOU, S., EL-BORAGY, M., GUST, W., and PREDEL, B. (1988) *Phase Transformations '87*, ed. G. W. Lorimer, p. 231, The Institute of Metals, London.
- ABE, T., MIYAZAKI, K., KAWAI, M., and HIRANO, K. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 359, Metals Society of AIME, Warrendale, PA.
- ACUÑA, R. and BONFIGLIOLI, A. (1974) *Acta Metall.* **22**, 399.
- AGARWAL, S. and HERMAN, H. (1973) *Acta Metall.* **21**, 1649.
- ALEXANDER, K. B., STEIGERWALD, D. A., and LAUGHLIN, D. E. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 945, Metals Society of AIME, Warrendale, PA.
- ARDELL, A. J. (1968) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 111, The Institute of Metals, London; (1972) *Acta Metall.* **20**, 61; (1988) *Phase Transformations '87*, ed. G. W. Lorimer, p. 485, The Institute of Metals, London.
- ARDELL, A. J. and NICHOLSON, R. B. (1966) *Acta Metall.* **14**, 1295.
- ASIMOV, R. (1963) *Acta Metall.* **11**, 72.
- ATKINSON, C. (1981) *Proc. R. Soc. London, Ser. A* **378**, 351; (1982) *Ibid.* **384**, 107.
- AUBAUER, H. F. (1978) *J. Phys.* **F 8**, 375.
- BECKER, R. (1937) *Z. Metallkd.* **29**, 245; (1938) *Ibid.* **32**, 128.
- BORELIUS, G. (1945) *Ark. Mat., Astron. Fys.* **32A**, No. 1; (1951) *Trans. Metall. Soc. AIME* **191**, 477; (1954) *Defects in Crystalline Solids*, p. 169, The Physical Society, London.
- BORELIUS, G. and LARSSON, L. E. (1956) *The Mechanism of Phase Transformations in Metals*, p. 67, The Institute of Metals, London.
- BORELIUS, G. and SÆSTEN, K. M. (1948) *Ark. Mat., Astron. Fys.* **36B**, No. 5.
- BORELIUS, G., LARISS, F., and OHISSON, E. (1944) *Ark. Mat., Astron. Fys.* **31A**, No. 10.
- BRADLEY, J. R., SHIFFLET, G. J. and AARONSON, H. I. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, 819 Metals Society of AIME, Warrendale, PA.
- BRADLEY, J. R., RIGSBEE, J. M., and AARONSON, H. I. (1977) *Metall. Trans.* **8A**, 323.
- BRAILSFORD, A. D. and WYNBLATT, P. (1979) *Acta Metall.* **27**, 489.
- BRANDRETH, M. and HELLAWELL, A. (1962) *Acta Metall.* **10**, 174.
- BUTLER, E. P. and WILLIAMS, D. B. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 909, Metals Society of AIME, Warrendale, PA.
- BUTLER, E. P., RAMASWAMY, V., and SWANN, P. R. (1973) *Acta Metall.* **21**, 517.
- CAHN, J. W. (1959) *Acta Metall.* **7**, 18; (1961) *Ibid.* **9**, 795; (1962) *Ibid.* **10**, 179; (1964) *Ibid.* **12**, 1457; (1965) *J. Chem. Phys.* **42**, 93.
- CAHN, J. W. and HILLIARD, J. E. (1958) *J. Chem. Phys.* **25**, 258.
- CALNAN, E. A. and CLEWS, C. J. B. (1952) *J. Iron Steel Inst.* **170**, 252.
- CALVERT, J., JACQUET, P., and GUINIER, A. (1939) *J. Inst. Met.* **65**, 121.
- CASTAING, R. (1956) *The Mechanism of Phase Transformations in Metals*, p. 283, The Institute of Metals, London.
- COATES, D. E. (1972) *Metall. Trans.* **3A**, 1203; (1973) *Ibid.* **4A**, 1077.
- COOK, H. E. (1970) *Acta Metall.* **18**, 297.
- COOK, H. E., and DE FONTAINE, D. (1969) *Acta Metall.* **17**, 915.
- COOK, H. E., DE FONTAINE, D., and HILLIARD, J. E. (1969) *Acta Metall.* **17**, 765; (1971) *Ibid.* **19**, 607.
- DANIEL, V. and LIPSON, H. (1943) *Proc. R. Soc. London, Ser. A* **181**, 368; (1944) *Ibid.* **182**, 378.
- DASH, W. C. (1956) *J. Appl. Phys.* **27**, 1193.
- DAVIES, C. K. L., NASH, P., and STEVENS, R. N. (1980) *Acta Metall.* **28**, 179.
- DEHLINGER, U. (1937) *Z. Metallkd.* **29**, 401.

- DEHOFF, R. T. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 503, Metals Society of AIME, Warrendale, PA.
- DE SORBO, W., TREAFTIS, H. N., and TURNBULL, D. (1958) *Acta Metall.* **6**, 401.
- DE SORBO, W. and TURNBULL, D. (1956) *Acta Metall.* **4**, 495; (1959) *Ibid.* **7**, 83.
- DOHERTY, R. D. and CANTOR, B. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 547, Metals Society of AIME, Warrendale, PA.
- DOI, M. (1996) *Prog. Mater. Sci.* **40**, 79.
- DOI, M., FUKAYA, M., and MIYAZAKI, T. (1996) *Philos. Mag.* **57A**, 821.
- DOI, M. and MIYAZAKI, T. (1992) *J. Mater. Sci.* **127**, 629.
- DOREMUS, R. H. (1957) *Acta Metall.* **5**, 393; (1958) *Ibid.* **6**, 674; (1959) *Ibid.* **7**, 399; (1960) *Trans. Metall. Soc. AIME* **218**, 596.
- DOREMUS, R. H. and KOCH, E. F. (1960) *Trans. Metall. Soc. AIME* **218**, 591.
- ENOMOTO, M. (1987) *Acta Metall.* **35**, 947.
- ENOMOTO, M., TOKUYAMA, M., and KAWASAKI, K. (1986) *Acta Metall.* **34**, 2119.
- ENOMOTO, M. and AARONSON, H. I. (1985) *Scr. Metall.* **19**, 1.
- FISHER, J. C. and HOLLOMON, J. H. (1955) *Acta Metall.* **3**, 608.
- FOURNELLE, R. A. and CLARK, J. B. (1972) *Metall. Trans.* **3**, 2757.
- FOURNELLE, R. A., JU, C. P., YANG, C. F., and SARKAR, G. (1988) *Phase Transformations '87*, ed. G. W. Lorimer, p. 238, The Institute of Metals, London.
- FRANK, F. C. and TURNBULL, D. (1956) *Phys. Rev.* **104**, 617.
- FREISE, E. J., KELLY, A., and NICHOLSON, R. B. (1961) *Acta Metall.* **9**, 250.
- GEISLER, A. H. and HILL, J. K. (1948) *Acta Crystallogr.* **1**, 238.
- GEROLD, V. (1954) *Z. Metallkd.* **45**, 599.
- GLOCKER, R. G., KOSTER, W., SCHERB, J., and ZIEGLER, G. (1952) *Z. Metallkd.* **43**, 208.
- GREENWOOD, G. W. (1968) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 103, The Institute of Metals, London.
- GUINIER, A. (1942) *J. Phys. Radium* **3**, 124; (1950) *C. R. Acad. Sci.* **206**, 1641; (1959) *Solid State Phys.* **9**, 293.
- GUST, W. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 921, Metals Society of AIME, Warrendale, PA.
- GUST, W., LEININGER, U., and PREDEL, B. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 927, Metals Society of AIME, Warrendale, PA.
- HAM, F. S. (1959a) *J. Appl. Phys.* **30**, 915; (1959b) *Ibid.* **30**, 1518.
- HARDY, H. K. and HEAL, T. J. (1954) *Prog. Met. Phys.* **5**, 1; (1956) *The Mechanism of Phase Transformations in Metals*, p. 7, The Institute of Metals, London.
- HART, E. W. (1958) *Acta Metall.* **6**, 553.
- HILLERT, M. (1961) *Acta Metall.* **9**, 525.
- HILLIARD, J. E. (1970) *Phase Transformations*, p. 497, American Society for Metals, Metals Park, OH.
- HIROTSU, Y., and NAGAKURA, S. (1972) *Acta Metall.* **20**, 645.
- HOBSTETTER, J. N. (1949) *Trans. Metall. Soc. AIME* **180**, 121.
- HOWE, J. M., DAHMEN, U., and GRONSKY, R. (1987) *Philos. Mag.* **56A**, 31.
- HUSTON, E. L., CAHN, J. W., and HILLIARD, J. E. (1966) *Acta Metall.* **14**, 1053.
- HYMAN, E. D. and NUTTING, J. (1956) *J. Iron Steel Inst.* **184**, 148.
- JACK, K. H. (1951) *Proc. R. Soc. London, Ser. A* **208**, 216.
- JAGODZINSKI, H. and LAVES, F. (1949) *Z. Metallkd.* **40**, 296.
- JONES, G. J. and TRIVEDI, R. K. (1971) *J. Appl. Phys.* **42**, 4299; (1975) *J. Cryst. Growth* **29**, 155.
- KAWASAKI, K. and TOKUYAMA, M. (1987) *Acta Metall.* **35**, 907.
- KHACHATURYAN, A. G., SEMENOVSKAYA, S. V., and MORRIS, J. M. (1988) *Acta Metall.* **36**, 1563.
- KING, A. D. and BELL, T. (1975) *Metall. Trans.* **6**, 1428.
- KINSMAN, K. R. and AARONSON, H. I. (1973) *Metall. Trans.* **4**, 959.
- KÖSTER, W. and BRAUMANN, F. (1952) *Z. Metallkd.* **43**, 193.
- KÖSTER, W. and KNODLER, A. (1955) *Z. Metallkd.* **46**, 632.

- KRAUSS, G. (1984) *Phase Transformations in Ferrous Alloys*, eds. A. H. Marder and J. L. Goldstein, p. 101, Metals Society of AIME, Warrendale, PA.
- LANGER, J. S. (1973) *Acta Metall.* **21**, 1649.
- LAIRD, C. and AARONSON, H. I. (1967) *Acta Metall.* **15**, 73.
- LANGE, W. F., III and AARONSON, H. I. (1979) *Metall. Trans.* **10A**, 1951.
- LEMENT, B. S., AVERBACH, B. L., and COHEN, M. (1954) *Trans. Am. Soc. Met.* **46**, 851.
- LESLIE, W. C. (1961) *Acta Metall.* **9**, 1004.
- LIFSHTZ, F. M. and SLYOZOV, J. (1961) *J. Phys. Chem. Solids* **19**, 35.
- LIVINGSTON, J. D. (1959) *Trans. Metall. Soc. AIME* **215**, 566.
- LOGAN, R. A. (1955) *Phys. Rev.* **100**, 615.
- MARQUESE, J. A. and ROSS, J. (1984) *J. Chem. Phys.* **80**, 536.
- MILLER, M. K., BEAVAN, P. A., BRENNER, S. S., and SMITH, G. D. W. (1983) *Metall. Trans.* **14A**, 1021.
- MILLER, M. K., HYDE, J. M., HETHERINGTON, M. G., CEREZO, A., SMITH, G. D. W., and ELLIOTT, C. M. (1995a) *Acta Metall. Mater.* **43**, 3403; (1995b) *Ibid.* **43**, 3415.
- MIZAZAKI, T., DOI, M., and KOZAKAI, T. (1988) *Solid State Phenomena* **3-4**, 227.
- MIZAZAKI, T., IMAMURA, H., and KOZAKAI, T. (1982) *Mater. Sci. Eng.* **54**, 9.
- MIZAZAKI, T., IMAMURA, H., MORI, H., and KOZAKAI, T. (1981) *J. Mater. Sci.* **16**, 1197.
- NAGAKURA, S., HIROTSU, Y., KUSUNOKI, M., SUZUKI, T., and NAKAMURA, Y. (1983) *Metall. Trans.* **14A**, 1025.
- NICHOLSON, R. B. and NUTTING, J. (1958) *Philos. Mag.* **3**, 531; (1961) *Acta Metall.* **9**, 332.
- NICHOLSON, R. B., THOMAS, G., and NUTTING, J. (1958-9) *J. Inst. Met.* **87**, 427.
- NYSTRÖM, J. (1949) *Ark. Fys.* **1**, 359.
- PANSERI, C., GATTO, F., and FEDERIGHI, T. (1958) *Acta Metall.* **6**, 198.
- PETERMANN, J. and HORNBOGEN, E. (1968) *Z. Metallkd.* **59**, 814.
- PITSCH, W. (1955) *Acta Metall.* **3**, 542; (1957) *Ibid.* **5**, 175.
- PITSCH, W. and LUCKE, K. (1956) *Arch. Eisenhuettenwes.* **27**, 45.
- PRESTON, G. D. (1938a) *Proc. R. Soc. London, Ser. A* **167**, 526; (1938b) *Philos. Mag.* **26**, 855.
- PURDY, G. R. and KIRKALDY, J. S. (1963) *Trans. Metall. Soc. AIME* **227**, 1235.
- REYNOLDS, W. T., ENOMOTO, J. R., and AARONSON, H. (1984) *Phase Transformations in Ferrous Alloys*, eds. A. R. Marder and J. L. Goldstein, p. 155, Metals Society of AIME, Warrendale, PA.
- ROSENBAUM, H. S. and TURNBULL, D. (1958) *Acta Metall.* **6**, 563.
- RUNDMAN, K. B. and HILLIARD, J. E. (1967) *Acta Metall.* **15**, 1025.
- SARIAN, S. and WEART, H. W. (1969) *J. Appl. Phys.* **37**, 1675.
- SCHEIL, E. (1952) *Z. Metallkd.* **43**, 40.
- SEITH, W. and LAIRD, J. G. (1932) *Z. Metallkd.* **24**, 193.
- SEITZ, F. (1952) *Adv. Phys.* **1**, 43.
- SILCOCK, J. M. and HEAL, T. J. (1956) *Acta Crystallogr.* **9**, 680.
- SILCOCK, J. M., HEAL, T. J., and HARDY, H. K. (1953-4) *J. Inst. Met.* **82**, 239.
- SMITH, C. S. (1952) *Imperfections in Nearly Perfect Crystals*, p. 377, Wiley, New York; (1953) *Trans. Am. Soc. Met.* **45**, 533.
- SUN, Y., LIU, S., and WANG, H. (1988) *Phase Transformations '87*, ed. G.W. Lorimer, p. 246, The Institute of Metals, London.
- THOMAS, G. and NUTTING, J. (1959) *Acta Metall.* **7**, 515.
- THOMAS, G. and WHELAN, M. J. (1959) *Philos. Mag.* **4**, 511.
- TIEDMA, T. J. and BURGERS, W. G. (1954) *Appl. Sci. Res., Sect. A* **4**, 243.
- TOKUYAMA, M. and KAWASAKI, K. (1984) *Physica* **123A**, 386.
- TSOU, A. L., NUTTING, J., and MENTOR, J. W. (1952) *J. Iron Steel Inst.* **172**, 163.
- TURNBULL, D. (1954) *Defects in Crystalline Solids*, p. 203, The Physical Society, London; (1955) *Acta Metall.* **3**, 55; (1956) *Solid State Phys.* **3**, 226.
- TURNBULL, D. and TREAFTIS, H. N. (1955) *Acta Metall.* **3**, 43; (1957) *Ibid.* **5**, 534; (1958) *Trans. Metall. Soc. AIME* **212**, 33.
- TWEET, A. G. (1957) *Phys. Rev.* **106**, 221.
- VAN DER VEN, A. and DELAEE, L. (1996) *Prog. Mater. Sci.* **40**, 181.
- VOORHEES, P. W. and GLICKSMAN, M. E. (1984) *Acta Metall.* **32**, 2001, 2013.

- WAGNER, C. (1961) *Z. Elektrochem.* **65**, 581.
- WALKER, C. B., BLIN, J., and GUINIER, A. (1952) *C. R. Acad. Sci.* **235**, 254.
- WALKER, C. B. and GUINIER, A. (1953) *Acta Metall.* **1**, 568.
- WATSON, J. D. and MCDUGALL, P. G. (1973) *Acta Metall.* **21**, 961.
- WEINS, J. J. and CAHN, J. W. (1973) *Mater. Sci. Res.* **6**, 151.
- WERT, C. (1949) *J. Appl. Phys.* **20**, 943; (1950) *Phys. Rev.* **79**, 601.
- WERT, C. and ZENER, C. (1950) *J. Appl. Phys.* **21**, 5.
- WILLIAMS, D. B. and BUTLER, E. P. (1981) *Int. Met. Rev.* **26**, 153.
- WOOD, M. S. and HELLAWELL, A. (1961) *Acta Metall.* **9**, 174.
- YOSHIDA, H. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 363, Metals Society of AIME, Warrendale, PA.
- ZIEGLER, G. (1952) *Z. Metallkd.* **43**, 213.

## CHAPTER 17

# *Eutectoidal Transformations*

### 77. EUTECTOIDAL DECOMPOSITION AND *T-T-T* DIAGRAMS

The decomposition of a solid phase ( $\gamma$ ) to give two new phases ( $\alpha$  and  $\beta$ ) is illustrated in Fig. 17.1. This is a eutectoidal reaction and, at the temperature  $T^E$  of the horizontal line, three phases are in equilibrium according to the equation

$$\gamma(x^\gamma) = \alpha(x^\alpha) + \beta(x^\beta) \quad (77.1)$$

The best-known eutectoidal reaction is the decomposition of austenite into ferrite + cementite in (metastable) iron-carbon alloys. In the whole field of transformation studies, this is probably the most important reaction from a commercial viewpoint.

An alloy of the eutectoidal composition will transform at temperatures below  $T^E$  in such a way that the  $\alpha$  and  $\beta$  phases are formed together. If there are no metastable transformation products, this must be a discontinuous reaction in which the transformed  $\alpha + \beta$  regions are clearly separated from the  $\gamma$  phase. In steels, the phases ferrite and cementite form a characteristic lamellar structure known as pearlite. The term pearlite is sometimes used for the products of other eutectoidal transformations in which the two phases are in the form of alternating fine lamellae.

During continuous cooling from the  $\gamma$  phase, alloys having  $x < x^\gamma$  will usually precipitate primary  $\alpha$  crystals, followed by the eutectoidal decomposition in which  $\alpha$  and  $\beta$  crystals are precipitated together. Alloys with  $x < x^\gamma$  are called hypoeutectoidal alloys; those with  $x > x^\gamma$  are termed hypereutectoidal. The primary or pro-eutectoid phase in hypoeutectoidal alloys is  $\alpha$  and in hypereutectoidal alloys is  $\beta$ . The temperatures at which pro-eutectoidal  $\alpha$  or  $\beta$  and eutectoidal decomposition begin depend on the cooling rate and are of great importance in practical heat treatment. Nevertheless, little understanding of the mechanisms of transformation may be obtained from continuous cooling experiments and progress in this field was slow until the pioneering work of Bain and his colleagues showed the importance of isothermal transformation studies.

In an isothermal reaction, after quenching from the  $\gamma$  phase, the typical form of the curve of percentage transformed  $\zeta$  against time  $t$  is sigmoidal, as in other nucleation and growth reactions. If these isothermals are determined at different temperatures, a *T-T-T*

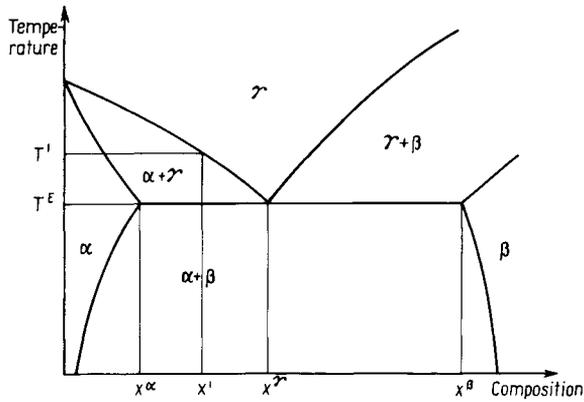
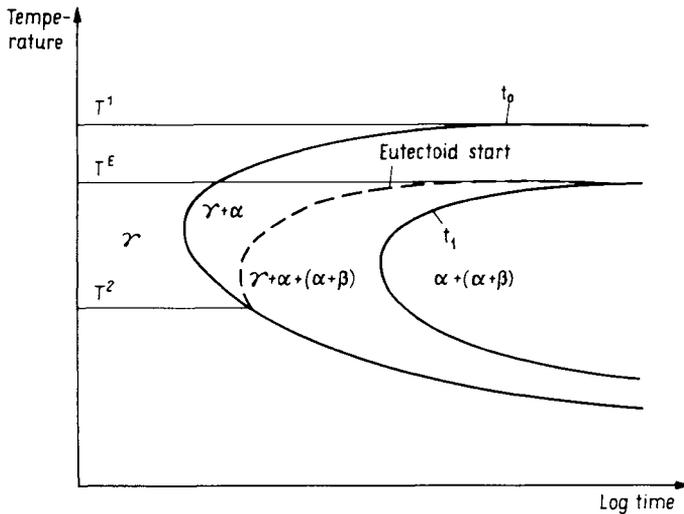


FIG. 17.1. Equilibrium diagram for a eutectoidal transformation.

FIG. 17.2. Hypothetical  $T-T-T$  diagram for the decomposition of the  $\gamma$  phase.

(time-temperature transformation) diagram may be drawn. The typical  $\log \zeta - T$  curve on a  $T-T-T$  diagram is C-shaped.

The isothermal transformation of a hypoeutectoidal alloy such as  $x^1$  in Fig. 17.1 will usually occur in two stages, the first being the separation of pro-eutectoid  $\alpha$  and the second the eutectoidal reaction itself. It is customary to include this information on the  $T-T-T$  diagram as shown in Fig. 17.2. Between the temperatures  $T^1$  and  $T^E$  only pro-eutectoid  $\alpha$  is formed. As the temperature is lowered below  $T^E$  the formation of the pro-eutectoid phase gradually diminishes and is replaced by the eutectoidal reaction, and eventually, below  $T^2$ , the whole product consists of the finely divided eutectoidal

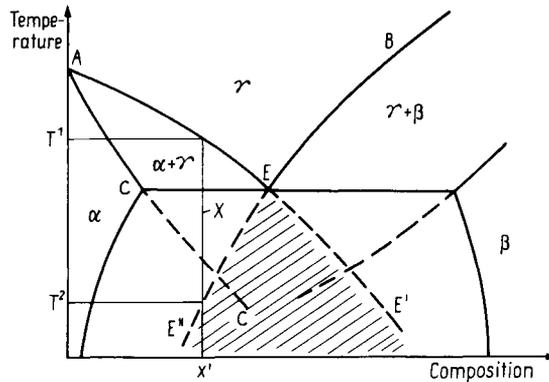


FIG. 17.3. The equilibrium diagram of Fig. 17.1 showing the extrapolated phase boundaries.

mixture. The C-curves divide the diagram into areas within which the products may be specified, as indicated in Fig. 17.2.

The significance of the temperature  $T^2$  may be illustrated by reference to Fig. 17.3. This shows the same equilibrium diagram as Fig. 17.1 but with the  $\alpha/\alpha + \gamma/\gamma$  and  $\beta/\beta + \gamma/\gamma$  boundaries extrapolated below the temperature  $T^1$ . The lines  $ACC'$  and  $AEE'$  may be interpreted as the equilibrium  $\alpha/\alpha + \gamma$  and  $\alpha + \gamma/\gamma$  phase boundaries in the absence of  $\beta$  nuclei and, in the hypothetical phase diagram without  $\beta$ , the transformation  $\gamma \rightarrow \gamma + \alpha$  would occur in an alloy of composition  $X^1$  below the temperature  $T^1$ . The same alloy is however unsaturated with respect to the hypothetical equilibrium diagram without  $\alpha$  and the transition  $\gamma \rightarrow \gamma + \beta$  would not begin until the temperature was further lowered to below  $T^2$ . The shaded area of the diagram shows the region in which alloys are supersaturated with respect to both  $\alpha$  and  $\beta$  and the hypothetical separate diagrams and, within this region, decomposition of  $\gamma$  into  $\alpha + \beta$  should be the first reaction. Alloys quenched to points such as  $X$  however will first deposit  $\alpha$  precipitate and the eutectoidal reaction will not begin until the decomposition of the residual  $\gamma$  has changed sufficiently for it to reach the extrapolated  $\gamma/\gamma + \beta$  boundary.

The formation of a pro-eutectoid phase is a simple precipitation reaction and need not be considered further in this chapter, which is concerned mainly with the factors governing the nucleation and growth of discontinuous cells of the transformation product. For simplicity, the assumption will generally be made that the alloy has the overall composition of the eutectoid,  $x^E$ . Before discussing the reaction in detail, it is useful to comment on the complex form of many  $T-T-T$  curves for eutectoidal decompositions.

Figures 17.4-17.6 show the  $T-T-T$  diagrams for three different alloy steels. In Fig. 17.4 there are two distinct sets of curves, the upper of which represent the times for various percentages of transformation in the pearlitic range, the time for the initiation of the reaction being shown by the broken line. The lower set of C-curves represents the quite different bainite reaction which is discussed separately in Chapter 25. There is a small temperature gap between the two sets of C-curves in Fig. 17.4 and a steel of this

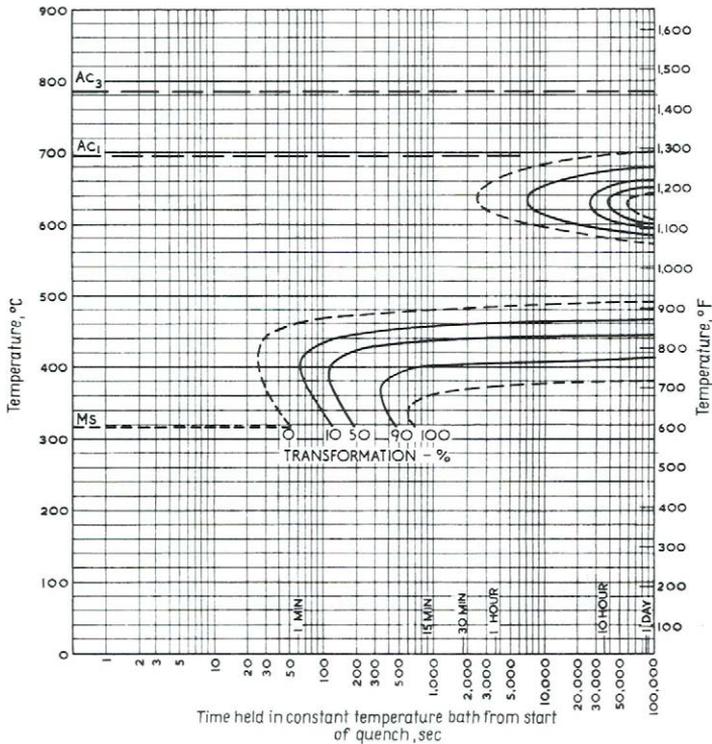


FIG. 17.4. Isothermal transformation diagram for EN-20-21/2 nickel chromium molybdenum steel (after B.I.S.R.A., 1956).

composition will remain almost indefinitely austenitic if cooled rapidly to a temperature around  $510^{\circ}\text{C}$ . The lower set of C-curves is drawn to terminate abruptly at the temperature  $M_s$ , where martensite begins to form; however, isothermal transformation to bainite may continue below  $M_s$  after the martensitic reaction has ended.

Figure 17.5 shows a similar set of C-curves but the transformations to bainite and to pearlite now overlap in the temperature range  $490\text{--}540^{\circ}\text{C}$ . Isothermal reaction in this temperature range results in the formation first of bainite and then of pearlite. The amount of overlap is still greater in steels of other compositions, until it becomes impossible to distinguish between two sets of C-curves. This is shown in Fig. 17.6, where one set of C-curves represents transitions to both kinds of product. Alloys of eutectoidal composition in plain carbon steels also have  $T\text{--}T\text{--}T$  diagrams in which only one family of C-curves can be distinguished; these alloys have bainitic structures when reacted at temperatures below the nose of the C-curves and pearlitic structures above the nose.

Various suggestions (see, e.g., Darken and Fisher, 1962; Shiflet, 1988) have been made that there are crystallographic orientation relations between the three phases, but other workers have found no systematic relation with the parent phase which is generally assumed to be randomly oriented with respect to the two product phases. This statement

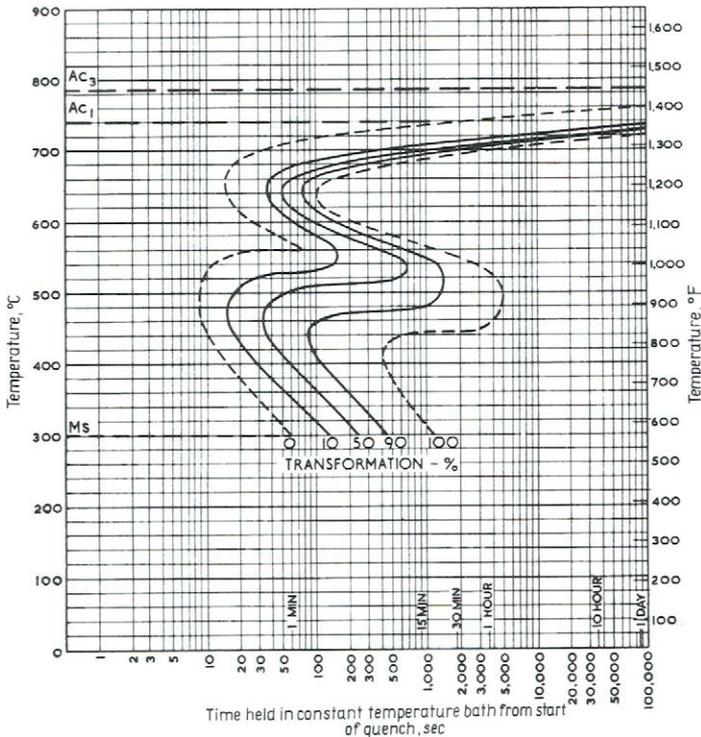


FIG. 17.5. Isothermal transformation diagram for EN-18/0.1% chromium steel (after B.I.S.R.A., 1956).

applies to the grain into which the lamellar product is growing; in order to obtain an interface of sufficiently low energy to permit nucleation, there must presumably be a systematic relation of some kind between the orientation of each product crystal and the initial grain into which the product did not grow.

In the high temperature region, it is possible to measure both growth rate and nucleation rate of the growing lamellar nodules. Various methods of measuring  $I$  were listed by Hull *et al.* (1942). The most direct method (Scheil and Lang-Weise, 1937) is rigorous for spherical nodules and consists of determining the number of nodules per unit volume in a series of specimens reacted for different times at the same temperature. The nucleation rate is then the time derivative of the number and the procedure is equivalent to that described on p. 17. This type of investigation is tedious and there are very few results described in the literature. The short-cut methods of Hull *et al.* involve rather questionable assumptions about growth rate and nodule shape, and in some cases they have been incorrectly applied. The best of these methods uses a statistical analysis of the distribution of sizes of the transformed regions in a single specimen, from which the variation of nucleation rate with time up to the total reaction time of the specimen may be evaluated. The procedure is valid only if all growing regions have the same growth rate which does

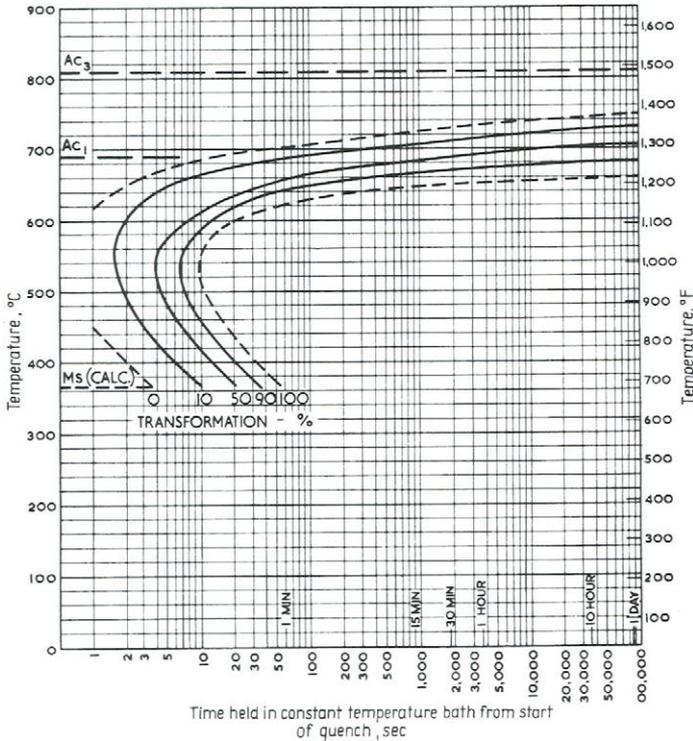


FIG. 17.6. Isothermal transformation diagram for EN-12.1% nickel steel (after B.I.S.R.A., 1956).

not change with time. Cahn and Hagel (1962) specified evidence against this assumption which they believed is responsible for some inconsistencies in reported nucleation rates. Other methods of determining the nucleation rate are still less reliable, being based on eqn. (4.7) with the assumption that  $I$  is constant.

Martensitic reactions take over from massive reactions at high cooling rates in other decompositions of eutectoidal phases and it seems that a high temperature phase which decomposes eutectoidally in equilibrium can rarely be quenched so as to retain its structure at lower temperatures. Other reactions occurring simultaneously with the eutectoidal decomposition may further complicate the  $T$ - $T$ - $T$  diagram. A suggested form of the  $T$ - $T$ - $T$  diagram for a Cu-12 at.% Al alloy is shown in Fig. 17.7 taken from the work of Haynes (1954-5).

## 78. THE AUSTENITE-PEARLITE TRANSFORMATION IN STEELS

The metastable iron-iron carbide equilibrium diagram is shown in Fig. 17.8. At a temperature of 723°C, a solid solution (austenite) containing 0.8 wt.% carbon is in

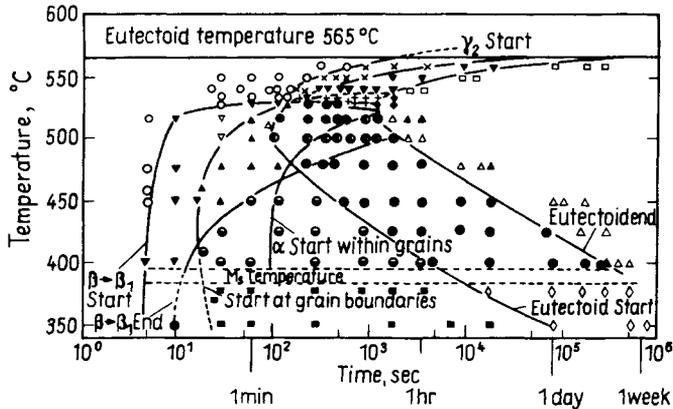


FIG. 17.7. Isothermal transformation diagram for Cu-12% Al alloy (after Haynes, 1954-5).

equilibrium with a solid solution (ferrite) containing about 0.025 wt.% carbon and the carbide phase  $\text{Fe}_3\text{C}$  (cementite). The characteristic microstructure to which the name pearlite was given has long been known to consist of alternate lamellae of ferrite and cementite. Pearlite usually nucleates on austenitic grain boundaries and grows as colonies which are regions within which the lamellae are parallel and the crystals of each type have a common orientation. A colony grows both by edgewise extension of the lamellae and sideways. The latter process probably does not require nucleation of new plates but takes place by branching from a single nucleus. This explains why all the plates of a colony have the same orientation, and was proved directly by Hillert (1962) who polished away a small pearlite colony and examined 340 successive sections. All the cementite lamellae were found to be branched from a common stem. New colonies are often observed to grow out from existing colonies, presumably as a result of a fresh nucleation event with the formation of a differently orientated nucleus.

At high temperatures, the nucleation rate is slow and the growth rate relatively rapid. Pearlite then grows from a small number of centres and takes the form of approximately spherical group nodules, each containing many colonies of parallel plates. The growing nodules do not seem to be hindered much by grain boundaries, and the growth rate is independent of grain size. Photographs showing the growth of pearlitic nodules through grain boundaries were obtained by Rathenau and Boas (1954) using an electron emission microscope. There is sometimes a slight change in lamellar orientation and it is surmised that the individual colonies may be halted at the boundary but that further colonies are nucleated almost immediately on the other side. The spreading group nodules continue to grow until they meet other nodules, and at high temperatures each such collection of colonies may grow to a size many times larger than the original grain size.

At low temperatures, nucleation is relatively easy and growth more difficult. The grain boundaries are then completely covered by pearlite colonies before the pearlite has penetrated far into the grains. Subsequent transformation requires mainly edgewise



There is some evidence from the formal kinetics (discussed in Section 56) that in this range nuclei form mainly at grain corners and not at random points on the grain boundaries.

The interlamellar spacing of pearlite depends strongly on the temperature at which it is produced. Very coarse lamellae are observed in group nodules growing just below the eutectoid temperature, and the carbide has some tendency to globularize. As the temperature is lowered, the spacing of the pearlite becomes progressively smaller and, at the lower end of the temperature range in which pearlite is formed, the individual plates can be resolved only in the electron microscope. Structures which were formerly called sorbite or nodular troostite because of the similarity of the microstructures to those found in tempered martensite have been shown by electron microscopy to be merely very fine pearlite.

Early studies (Mehl and Smith, 1935; Smith and Mehl, 1942) showed that the ferrite crystals in pearlite have an orientation relative to the austenite grain into which they are growing which is quite different from that of pro-eutectoid ferrite in hypoeutectoidal steels. Under the microscope, pro-eutectoid cementite often appears to be continuous with the cementite phase in pearlite, and Modin (1951) found an identity of orientation using polarized light. However, Modin also found cementite in pearlite to be continuous with pro-eutectoid cementite only when the pearlite was very fine. Nicholson (1954) used arguments based on observations such as the effect of carbon on the  $T-T-T$  diagrams of hypoeutectoid steels containing manganese. Increasing carbon content leads to the displacement to longer times of the curves representing the beginning of the formation of both ferrite and pearlite, suggesting that the nuclei for both reactions are similar. Dippenaar and Honeycombe (1973) studied morphological and crystallographic features of pearlite formation by transmission electron microscopy. They used a hypereutectoidal alloy containing 13 wt.% manganese and 0.8 wt.% carbon in which, after partial transformation to pearlite, the austenite could be retained to room temperature without forming martensite. This enabled the measurement of orientation relations with the parent austenite phase to be made. It was found that both pearlite phases were rationally related to the austenite grain on which they had formed but into which they did not grow. The ferrite had the Kurdjumov-Sachs orientation relation to the austenite and the cementite had the Pitsch orientation relation. This meant that the relation between the ferrite and the cementite was the Pitsch Petch orientation relation in which

$$[001]_c // [\bar{5}2\bar{1}]_f$$

$$[010]_c \text{ } 2\text{-}3^\circ \text{ from } [11\bar{3}]_f$$

$$[100]_c \text{ } 2\text{-}3^\circ \text{ from } [131]_f$$

This relation has been found by many other workers, but the alternative Bagaryatski-orientation relation in which

$$[100]_c // [0\bar{1}1]_f$$

$$[010]_c // [1\bar{1}\bar{1}]_f$$



is also often found and may even exist in the same specimen as the Pitsch–Petch relation. Dippenaar and Honeycombe suggest that the Pitsch–Petch relation is formed when both phases nucleate on a clean grain boundary, and the Bagaryatski relation occurs when the ferrite nucleates on the previously formed cementite.

Slightly different results were obtained by Fridberg and Hillert (1970) for pearlite nucleated on pro-eutectoid cementite with which it could be seen to be continuous. The orientation of the ferrite was determined entirely by the cementite with which it had a Bagaryatski relation. It is possible that the Pitsch–Petch relationship dominates at compositions close to that of the eutectoid whilst the Bagaryatski relation becomes increasingly important as the composition differs from that of the eutectoid point.

In a plain carbon steel of eutectoid composition, the cementite constitutes only about 12% by volume of the whole assembly. The question then arises why does the product have a lamellar rather than a needle-shaped form? One possible answer is that the surface free energy is very anisotropic.

Some results for the nucleation rate obtained by the first two methods described above are shown in Fig. 17.9 and indicate that the nucleation rate increases rapidly with time. In the early work of Hull *et al.* (1942) and Scheil and Lang-Weise (1937), the nucleation rate was found to vary approximately as the time squared. The rate is structure-sensitive and in particular is dependent on whether or not the specimen has been fully austenitized before cooling to the pearlitic transformation range. There has been much discussion as to

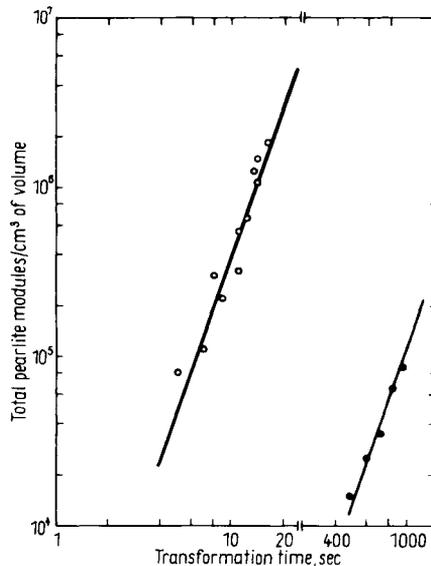


FIG. 17.9. Curves for total number of pearlite nodules versus time for two carbon steels (after Cahn and Hagel, 1962). Open circles: 1.01% C at 600 C (data of Hull *et al.*, 1942); filled circles: 0.93% C at 690 C (data of Scheil and Lang-Weise, 1937). The slope of the log-log plot is three, indicating that the nucleation rate increases as  $t^3$ .

whether ferrite or cementite is the active nucleus. Nucleation site saturation occurs readily except at high temperatures, after which the nucleation rate is, of course, zero and kinetic measurements can give no more information about nucleation.

The kinetics have hitherto been interpreted mainly in terms of a constant nucleation rate or a nucleation rate which decreases because of site saturation. Two explanations have been given for the increasing nucleation rate of pearlite, namely the theory of transient nucleation (Section 50) and the notion that nucleation is a two-stage process. The main assumption of the transient theory (Zeldovich, 1942; Turnbull, 1948) is that the time taken to establish a quasi-steady-state distribution of embryos is comparable with the whole reaction time. Turnbull's curves for the nucleation rate (Fig. 10.8) are of the same form as the observed effects.

The second explanation, due to Fisher (1950), is in terms of a two-stage nucleation process. Although this does not affect the subsequent argument, suppose for definiteness that the first nucleus is cementite and that the ferrite subsequently nucleates on the cementite-austenite interface. The formation of this double nucleus may be regarded as the effective starting event for pearlite. Suppose the first nucleus is cementite and that it grows parabolically according to the equation [see eqn. (54.10)]

$$r^2 = \alpha D^{1/2} t^{1/2} \quad (78.1)$$

If the growth is spherical, the interfacial area is

$$o = 4\pi\alpha^2 Dt$$

Now let the rate of nucleation of ferrite at the austenite-cementite interface be  $I''$  per unit area per unit time, and let  $p$  be the probability that the first ferrite nucleus has formed on the cementite particle. In a time interval  $dt$

$$\begin{aligned} dp &= (1-p)oI''dt \\ &= (1-p)4\pi\alpha^2 DI''tdt \end{aligned}$$

giving

$$p = 1 - \exp(-mt^2) \quad (78.2)$$

where  $m = 2\pi I''\alpha^2 D$ . The rate of nucleation per cementite particle is thus

$$I_{\alpha+\beta} = dp/dt = 2mt \exp(-mt^2) \quad (78.3)$$

This is related to  $I_\beta$  the rate of nucleation of cementite per unit area of austenite grain boundary; the number of such nuclei forming in the time interval  $+d\tau$  is  $I_\beta d\tau$  per unit area and the rate of nucleation of pearlite at these particles is

$$dI_{\alpha+\beta} = [2m(t-\tau) \exp\{-m(t-\tau)^2\} I''] d\tau \quad (78.4)$$

Assuming there are no nuclei present at time  $t = 0$ , the pearlitic nucleation rate after time  $t$  is obtained by integrating this equation from  $\tau = 0$  to  $\tau = t$ , to give

$$I_{\alpha+\beta} = I_\beta [1 - \exp(-mt^2)] \quad (78.5)$$

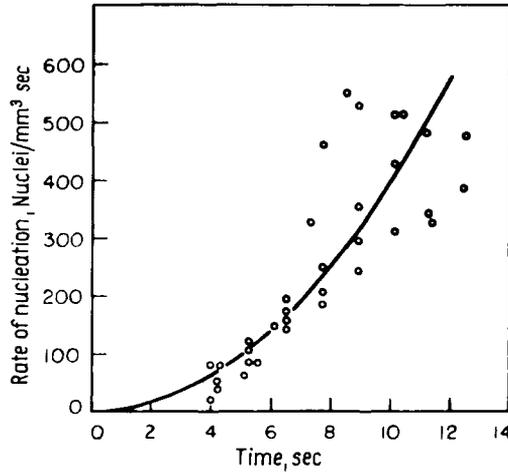


FIG. 17.10. Rate of nucleation of pearlite as a function of time (after Fisher, 1950). Full curve calculated from eqn. (78.6); experimental points from data of Hull *et al.* (1942).

where  $I_{\beta}$  is the nucleation rate of cementite per unit area of the  $\gamma$ - $\gamma$  grain boundary. At small values of  $mt^2$ , this relation gives

$$I_{\alpha+\beta} = I_{\beta}mt^2 \quad (78.6)$$

and at large values of  $mt^2$

$$I_{\alpha+\beta} = I_{\beta} \quad (78.7)$$

Figure 17.10 shows experimental points for the nucleation of a eutectoid steel at 680 °C, compared with a curve satisfying eqn. (78.6). Thus, according to this analysis, constant rates of nucleation per unit area of austenite grain boundary and of a growing interphase interface can lead to a steeply rising nucleation rate of pearlite with time, as is observed. Such a variation might be expected in all eutectoidal decompositions, but not in discontinuous precipitation where the  $\beta$  nucleus is also the nucleus for the  $\alpha + \beta$  aggregate. Clearly the same analysis applies if the primary nucleus is  $\alpha$  and the  $\beta$  phase is nucleated on the  $\alpha$ -parent phase interface.

Fisher's analysis rests on the assumptions of (i) constant nucleation rate of carbide particles, (ii) parabolic growth of these particles and (iii) nucleation rate of ferrite proportional to the austenite-cementite interfacial area. In addition, the predicted variation of nucleation rate with time will be valid only when eqn. (78.2) is valid, i.e. when only a few cementite particles have become pearlite nuclei. Cahn (1957) considered the effects of generalizing these assumptions. The first of them is probably incorrect at higher temperatures where, as already mentioned, the nuclei probably form on grain corners. The assumption of diffusion-controlled growth is reasonable for particles which have been growing for some time, so that the theory of Section 54 applies. In the early stages, however, it is possible that the cementite growth may be more nearly linear with time. The third assumption may be much lower at edges and corners where the

cementite–austenite interface meets the austenite grain boundary on which the cementite nucleated. Some support for this conclusion comes from the observation that pearlite does not seem to nucleate very readily on cementite particles in grain interiors.

A repetition of Fisher's derivation with more general assumptions leads to the conclusion that the nucleation rate will be proportional to the  $n$ th power of the time where

$$n = n_{\beta} + i(1 + n_{\zeta}) \quad (78.8)$$

In this equation,  $n_{\beta}$  is the exponent of time in the number of  $\beta$  carbide particles in existence during the transformation, being zero if all the carbide particles are present initially, one for constant nucleation rate and possibly greater than one if the nucleation increases through a transient stage, as in Turnbull's theory. As in eqn. (57.10)  $i$  is the dimensionality of the site on which ferrite nucleates on cementite (i.e.  $i = -2, 1$  and  $0$  for boundary, edge or corner nucleation respectively) and  $n_{\zeta} = 0$  for constant growth rate and  $-1/2$  for parabolic growth rate. If there is no transient in  $I_{\beta}$ , the maximum value of  $n$  in eqn. (78.8) is three, this being achieved by supposing the cementite to have constant nucleation rate and linear growth and that ferrite nucleates anywhere on the cementite–austenite interface.

Analysis of experimental results on steels containing carbides shows that they are often consistent with a two-stage model; the carbides do not themselves constitute pearlitic nuclei but may be converted into them. An analysis by Cahn (1957) however indicated that the nucleation rate may vary as  $t^3$  or  $t^4$ , and it is difficult to account for this variation unless part of it arises from a real transient in the nucleation rate of either cementite or ferrite.

The measured growth rate in a pearlitic reaction is usually obtained by plotting the diameter of the largest nodule visible in a microsection against the time. For all cases investigated, this gives a straight line, the slope of which is the constant velocity  $Y$ . This measured growth rate is clearly an overall average of the edgewise growth rate of a colony and the nucleation rate of new colonies within a nodule. The measured rate, however, is thought not to be very different from the true edgewise growth rate of an individual colony, especially as separate measurements of edgewise and sideways growth of some individual colonies have shown these rates to be nearly equal. The nodule diameter versus time relation does not pass through the origin when extrapolated back to zero size but has an intercept indicating zero size at a finite time. It is possible that this is a reflection of the variation of nucleation rate with time, the very low initial rate indicating that the chance of a nucleus being formed before a finite time has elapsed is virtually zero. Interpretations of this kind however must be made with extreme caution; the linear extrapolation may be quite unjustified and the growth rate may be very slow at first and then increase to a constant value.

The measurements show that the growth rate is not affected by structural irregularities due to incomplete austenitizing. However, some alloying elements produce large changes in growth rates, and when these are present the growth may be very dependent on the prior austenitizing treatment which affects the amount of solute in solution. Pronounced changes in growth rate are produced by molybdenum and smaller, but readily measured, changes by nickel and manganese; in all cases, the growth rate decreases with increasing amounts of these solutes. In the temperature range 600–700°C, addition of 0.5 at.%

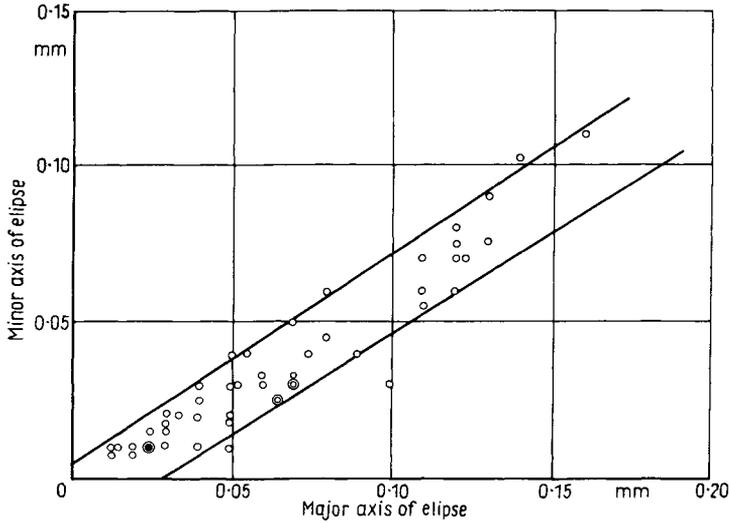


FIG. 17.11. Relation between edgewise and sideways growth of individual pearlite colonies in iron-molybdenum-carbon alloys (after Parcel and Mehl, 1952). The major axis is along the length of the pearlite plates.

molybdenum decreases the growth rate by a factor of more than 100. With the slow growth found in iron-carbon-molybdenum alloys, it is possible to measure individual colonies. The results obtained by Parcel and Mehl (1952) are shown in Fig. 17.11; they indicate that edgewise growth is about 1.4 times faster than sideways growth. The growth rate also increases with decreasing temperature below the eutectoid temperature, but not as rapidly as the nucleation rate.

Tests of the growth theories outlined in Chapter 11 depend on measurements of growth velocity, interlamellar spacing and temperature. New measurements on high purity alloys have replaced those summarized in the first edition, and have been reviewed by Ridley (1984). The quantity usually measured is the minimum observed interlamellar spacing but, because of variations in the true spacing, this differs from the true mean spacing. Measurements of minimum interlamellar spacing are generally consistent with Zener's prediction that  $y^{\alpha\beta}$  is proportional to  $1/\Delta T$ , where  $\Delta T$  is the supercooling  $T^E - T$  from the eutectoid temperature. Some measurements by Bolling and Richman (1970) showed deviations from this relation, but are ascribed by Ridley to recalescence. The isothermal measurements of Brown and Ridley (1969) and Cheetham and Ridley (1973) are in substantial agreement with those of Verhoeven and Pearson (1980) who used a "forced-velocity" technique in which the specimen is held in a moving temperature gradient. These results indicated a value for  $y^{\alpha\beta}\Delta T$  of  $6.19\ \mu\text{m}^\circ\text{C}$  and the best line extrapolated to the correct  $T^E$ . In a survey of the various results obtained up to the date of their review article, Marder and Bramfitt (1975) selected as the best available value  $y^{\alpha\beta}\Delta T = 8.02\ \mu\text{m}^\circ\text{C}$ . These values are appreciably smaller than the earlier results obtained on commercial purity iron-carbon alloys.

The interfacial free energy of  $\sigma^{\alpha\beta}$  may be now used to test the theory of growth; its apparent value may be obtained from the experimental  $y^{\alpha\beta}\Delta T$  by combining eqn. (55.1) with an optimizing condition, e.g. maximum growth velocity or maximum rate of entropy production. In this way Ridley obtained values for  $\sigma^{\alpha\beta}$  of  $940 \text{ mJ m}^{-2}$  for the maximum growth rate assumption and  $630 \text{ mJ m}^{-2}$  for the assumption of minimum entropy production. Both values are higher than expected and this may indicate that neither optimization principle is operative.

Comparison of the measured growth rate with the theoretical value [eqn. (55.20)] using experimental values of  $D$  has been attempted by many authors. The large diffusivity of carbon atoms, even at temperatures below room temperature, makes it probable that, in plain carbon steels, the rate-controlling process is diffusion of carbon through the bulk volume. However, in an alloy steel where co-precipitation of a carbide other than cementite takes place, the necessary transfer of the solute metal is probably confined to the interface region which provides an efficient "short circuit" for the necessary transfer of solute. In an early review, Cahn and Hagel (1962) combined the maximum value of  $\alpha^m$ , obtained from the equilibrium diagram as explained in Section 55, with experimental results for  $T$  and  $y^{\alpha\beta}$  to give an apparent diffusion coefficient  $D_{\text{app}}$ . [If the curvature of the interface is taken into account, the value of  $\alpha^m$  is reduced by a factor  $1 - (y_{\text{min}}^{\alpha\beta}/y^{\alpha\beta})$  but all the indications are that this factor is greater than 0.8, as opposed to Zener's prediction of 0.5.] The data examined by Cahn and Hagel gave values of  $D_{\text{app}}/D$ , the ratio of the volume diffusion coefficient needed to give the observed growth rates to the actual diffusion coefficient. For the non-ferrous alloys and many of the alloy steels, this ratio is very large, indicating that eqn. (55.14) cannot be correct and that diffusion must mainly be along the interface. Volume diffusion is possible when the ratio is near to unity, as it is for some iron-carbon-manganese alloys, but it is somewhat surprising for iron-carbon alloys. Estimates of the width of the diffusion zone ahead of the interface also indicated in most cases that diffusion through the lattice was not the rate-controlling process.

More recent experiments have reduced the discrepancy between calculated and measured growth rates for iron-carbon alloys and have generally been interpreted as being not inconsistent with volume diffusion of carbon as the main mechanism of pearlitic growth. Figure 17.12 compares measured growth rates of Brown and Ridley (1969) and Frye *et al.* (1942) with growth rates calculated on the arbitrary assumption that  $y^{\alpha\beta} = 3y_{\text{min}}^{\alpha\beta}$ . Puls and Kirkaldy (1972) and Cheetham and Ridley (1973) have shown that, by taking the carbon diffusion coefficient at an appropriate average carbon content, the discrepancy between theory and experiment which was formerly thought to be as large as a factor of 50 is reduced to two to three times. It remains true, however, that the experimental velocities are higher than the theoretical velocities and it has been suggested that this is because of a contribution from interfacial diffusion.

The overall kinetics of pearlite formation depend on whether or not nucleation-site saturation is attained early in the transformation. When site saturation occurs,  $Yt_{1/2}/L^B \cong 0.1$  [see eqn. (57.8)] and Cahn (1957) used an approximate condition

$$Yt_{1/2}/L^B < 0.5 \quad (78.9)$$

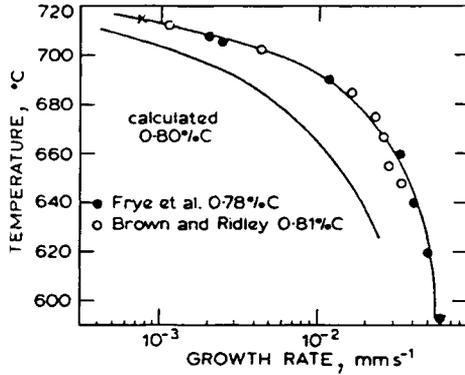


FIG. 17.12. Measured and calculated growth rates for Fe-C alloys. The calculated rate assumed that  $y^{\alpha\beta} = 3y_{\min}^{\alpha\beta}$ .

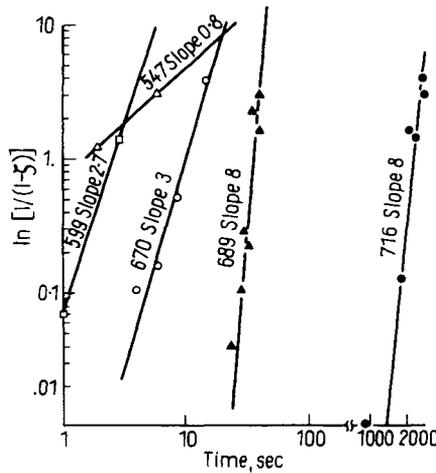


FIG. 17.13. Plot of  $\log \ln [1/(1 - \zeta)]$  versus  $\log t$  for pearlite formation in high purity slightly hypereutectoid steel (after Cahn, 1957; data of Hull *et al.*, 1942). The figures give the temperatures in degree Celsius.

for the onset of saturation. The data of Hull *et al.* then show that, in hypereutectoid steels, site saturation occurs at all temperatures investigated, and in all other steels the sites become saturated below about 660°C. When saturation occurs, the half-time depends only on the growth rate  $Y$  and attempts to deduce the nucleation rate are erroneous.

As noted above, the transition from unsaturated to saturated kinetic behaviour should take place over a narrow temperature interval and this transition is revealed as a change in the slope of experimental  $\log \ln [1/(1 - \zeta)]$  versus  $\log t$  curves, as the transformation equation approximates to eqns. (57.2), (57.5) or (57.6) after saturation (depending on the type of site) and to the equation for randomly distributed nuclei before saturation. Because of the time dependence of the nucleation rate in pearlite formation, the

unsaturated rate law is given by an equation of the form of eqn. (56.3) rather than by eqn. (57.1), but with  ${}^vN_0=0$ . Figure 17.13 shows experimental results for a steel of near-eutectoid composition; the slope is about eight at high temperatures and changes to three below 660°C and to one below 600°C.

These results confirm that site saturation begins at about 660°C. The high temperature slope suggests a nucleation rate proportional to the fourth power of the time, as mentioned above. In the temperature range 600–660°C, the effective nucleation sites are apparently grain corners [eqn. (57.6)] and general grain boundary nucleation only becomes effective below 600°C [eqn. (57.2)]. This is in agreement with the observed morphology. For pearlite formation in alloys of some other compositions, there is considerable scatter in the reported slopes; therefore this kind of analysis cannot be applied readily but the transition from unsaturated- to saturated-site behaviour is detectable in most cases.

### 79. THE FORMATION OF AUSTENITE FROM TWO-PHASE MIXTURES OF FERRITE AND CEMENTITE

When a mixture of the  $\alpha$  and  $\beta$  phases is heated to temperatures above a eutectoidal horizontal, a reaction takes place and the  $\gamma$  phase is formed. The kinetics of this kind of reaction have been mainly studied in the case of ferrite and cementite reacting to form austenite (Roberts and Mehl, 1943; Molinder, 1956). The reaction is discontinuous in the sense used in this chapter, and each region of transformation product is nominally a single crystal. The qualification is necessary because the experiments show that the growing  $\gamma$  crystals contain small undissolved carbide particles which disappear only after further holding at the reaction temperature.

For the cooling transformation, the only important structural parameter is the austenitic grain size, and its influence is confined mainly to the nucleation rate. In the reverse transformation, the type of carbide particle, its shape and state of dispersion are capable of exerting a considerable influence, not only on the nucleation rate but also on the growth rate. Nucleation takes place at the interface between ferrite and carbide and the number of potential sites of this kind is so large that only a few are actually utilized in forming austenite. Nevertheless, the rate of nucleation increases with the number of sites, i.e. with the ferrite–carbide interfacial area.

The diffusion process which is responsible for the growth of the austenite is not certain; it could be diffusion through one of the phases or, in the case of lamellar eutectoidal structures, along the incoherent interface in a reversal of the growth mechanism on cooling. In all cases, the mean diffusion distance will be proportional to the distance between carbide particles so that the growth rate should increase as this distance decreases. This is confirmed by results on the formation of austenite from pearlite; the growth rate being considerably greater for fine pearlite than for coarse pearlite.

When the carbides are present in the original assembly as pearlite lamellae, or as spheroidal particles resulting from transformation at high temperatures, the austenite grains are substantially spherical in shape. Nucleation sites are so numerous that site saturation does not occur; the kinetics of the reaction correspond to random volume

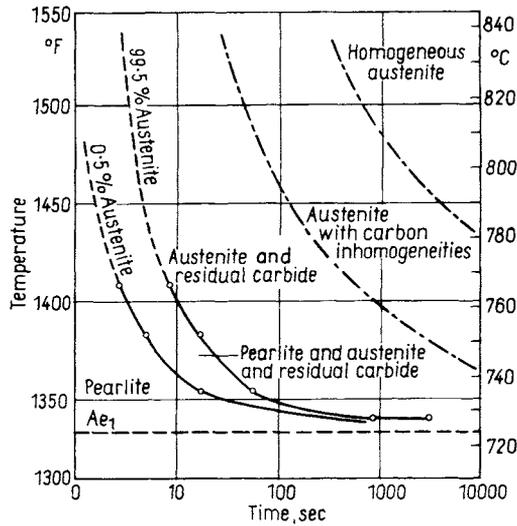


FIG. 17.14. Isothermal transformation diagram for formation of austenite from pearlite in a simple carbon eutectoidal steel (after Roberts and Mehl, 1943).

nucleation. Because the growth rate increases with increasing temperature above  $T^E$ , the  $T$ - $T$ - $T$  diagram does not consist of C-shaped curves, but shows an increasing reaction rate with increase in temperature. Figure 17.14 shows such a diagram for a eutectoidal steel. It will be noted that after transformation to austenite is virtually complete (as measured by the complete impingement of product regions and the disappearance of ferrite) there are still some undissolved carbides present. The two later stages of the reaction correspond to the disappearance of these particles and, finally, to the removal of carbon composition differences within the austenite solid solution. The whole action is strictly only completed when homogeneous austenite has been produced, but the discontinuous change ends with the formation of austenite residual carbide, and the carbides disappear by a slower continuous process.

Isothermal transformation curves from pearlite to austenite have the usual sigmoidal shape. As noted above, the reaction is faster the smaller the interlamellar spacing of the pearlite, and alloying elements also considerably influence the kinetics. An aluminium de-oxidized steel was found by Roberts and Mehl to give a much slower transformation rate than an untreated steel of the same composition. Transformation is also more rapid in quenched and tempered steels than in pearlitic and spheroidal steels, as would be expected from the much finer state of dispersion of the carbides.

Roberts and Mehl analysed the kinetics by finding  $Y$  from observations on individually growing grains and deducing  $t$  from eqn. (4.7). This procedure should be correct if the kinetics correspond to constant nucleation rate and three-dimensional growth. Table XV shows results for growth and nucleation rates for one of Roberts and Mehl's steels. The number of nuclei formed per unit volume apparently decreases with temperature, although

TABLE XV. MEASURED GROWTH AND NUCLEATION RATES FOR THE FORMATION OF AUSTENITE FROM A FINE PEARLITIC STEEL (ASSUMING EQUATION (79.1))

Temperature (°C)	Linear Growth Rate (mm sec <sup>-1</sup> )	Nucleation Rate (mm <sup>-3</sup> sec <sup>-1</sup> )
727	$3.2 \times 10^{-5}$	$1.9 \times 10^{-2}$
735	$8.9 \times 10^{-4}$	$3.5 \times 10^{-3}$
746	$2.5 \times 10^{-3}$	$1.0 \times 10^{-4}$

the driving force is increasing. This may be a result of a transient in the nucleation rate, so that the higher growth rate at higher temperatures leads to possible nucleation sites being absorbed by growing crystals before they can be activated.

The formation of austenite from quenched steels or quenched and tempered steels is apparently a more complex process. The austenite regions in such assemblies grow as plates rather than spheres (Nehrenberg, 1950). The evidence for the retention of some carbide after the discontinuous reaction is complete is mainly metallographic, and selective etching techniques also show the presence of inhomogeneous regions of austenite before the carbon is uniformly distributed in solution at a later stage. Roberts and Mehl attempted to obtain quantitative data on the carbon heterogeneities by measuring the nucleation rate of pearlite in incompletely austenitized specimens. Unfortunately, there are two independent effects leading to a decrease in nucleation rate per unit volume as the austenitizing time is increased. One of these is the increase in austenite grain size and the other the removal of carbide particles. Cahn (1957) showed that the original way of separating these effects was probably faulty, as it ignored the fact that the transformations to pearlite were made near the temperature range at which nucleation-site saturation takes place. Some of the results can be explained entirely by the variation in austenite grain size; in others, as already described, there is an increase in pearlite nucleation rate because of the presence of carbides or carbon-rich regions of austenite.

The initial structure in the work of Roberts and Mehl was mainly lamellar pearlite. A detailed investigation of austenitization in an annealed 1.27% carbon alloy having an initial structure of rounded cementite grains in a ferritic matrix was made by Molinder (1956). He found that the austenite regions forming at the beginning of transformation had a low carbon content, but that this increases progressively as cementite dissolves, so that the carbon content of the austenite when the last traces of ferrite disappeared was 0.55–0.6% at all the temperatures he investigated.

Molinder distinguished three stages of the reaction. The first is an incubation or slow nucleation period before austenite can be detected, and this was found only at 750°C, the lowest temperature investigated. The second stage is the disappearance of the remaining cementite and the formation of austenite of equilibrium composition. Figure 17.15 shows the carbon content of the austenite as a function of time at different temperatures. The point corresponding to the disappearance of ferrite is marked on the curves, and it

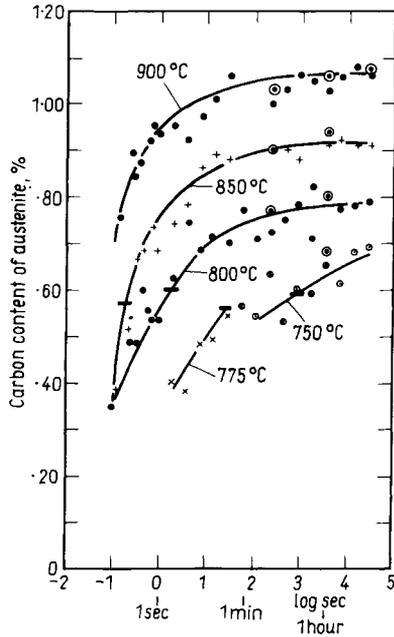


FIG. 17.15. The carbon content of austenite versus the logarithm of the time at austenitizing temperature (after Molinder, 1956).

seems that the dissolution of cementite is a continuous process before and after the transformation to austenite is completed.

For this type of initial structure, the first or discontinuous part of the reaction is essentially a lattice transformation of ferrite to austenite, followed by the continuous solution of the carbides in the austenite. In lamellar pearlitic steels, it seems rather more likely that the formation of the austenite involves reaction between the ferrite and cementite, but even in these circumstances the carbide does not all dissolve. Mehl and Hagel (1956) estimated that the activation energy for the final stages of austenization is consistent with the assumption that the diffusion of carbon in austenite is the rate-determining process. Molinder, however, found an activation energy of approximately  $800 \text{ kJ mole}^{-1}$ , which is about six times the known activation energy for diffusion of carbon. He therefore suggested that the rate of disintegration of cementite at the interface is the controlling factor.

Gupta and Nakkalil (1988) have investigated the dissolution of the cellular precipitate in a copper-cadmium alloy. They found that dissolution began from the original position of the grain boundary and the cellular boundary, and the microstructure during dissolution resembled that during precipitation. Applying a theory due to Petermann and Hornbogen (1968), they deduced that the activation energy for the dissolution is  $137 \text{ kJ mole}^{-1}$ , which is about three-quarters of the activation energy for volume diffusion, and the authors suggest that the rate-determining process is diffusion along grain boundaries.

## REFERENCES

- B.I.S.R.A. (1956) *Atlas of Isothermal Transformations Diagrams*, Iron and Steel Institute, London (Special Report No. 56).
- BOLLING, G. F. and RICHMAN, R. H. (1970) *Metall. Trans.* **1**, 2095.
- BROWN, D. and RIDLEY, N. (1969) *J. Iron Steel Inst.* **207**, 1232.
- CAHN, J. W. (1957) *Trans. Metall. Soc. AIME* **209**, 140.
- CAHN, J. W. and HAGEL, W. C. (1962) *Decomposition of Austenite by Diffusional Processes*, p. 131, Interscience, New York.
- CHEETHAM, D. and RIDLEY, N. (1973) *J. Iron Steel Inst.* **211**, 648.
- DARKEN, L. S. and FISHER, R. M. (1962) *Decomposition of Austenite by Diffusional Processes*, p. 249, Interscience, New York.
- DIPPENAAR, R. J. and HONEYCOMBE, R. W. K. (1973) *Proc. R. Soc. London, Ser. A* **333**, 433.
- FISHER, J. C. (1950) *Thermodynamics in Physical Metallurgy*, p. 201, American Society for Metals, Cleveland, OH.
- FRIDBERG, J. and HILLERT, M. (1970) *Acta Metall.* **18**, 1253.
- FRYE, J. H., SANSBURY, E. E., and McELROY (1942) *Trans. Metall. Soc. AIME* **30**, 183.
- GUPTA, S. P. and NAKKALIL, R. (1988) *Phase Transformations-87*, ed. G. W. Lorimer, p. 269, The Institute of Metals, London.
- HANSEN, M. (1958) *The Constitution of Binary Alloys*, McGraw-Hill, New York.
- HAYNES, R. (1954-5) *J. Inst. Met.* **83**, 105.
- HILLERT, M. (1962) *Decomposition of Austenite by Diffusional Processes*, p. 197, Interscience, New York.
- HULL, F. C., COLTON, R. A., and MEHL, R. F. (1942) *Trans. Metall. Soc. AIME* **150**, 185.
- MARDER, A. R. and BRAMFITT, B. L. (1975) *Metall. Trans.* **7A**, 902.
- MEHL, R. F. and HAGEL, W. C. (1956) *Prog. Met. Phys.* **6**, 74.
- MEHL, R. F. and SMITH, D. W. (1935) *Trans. Metall. Soc. AIME* **116**, 330.
- MOLINDER, G. (1956) *Acta Metall.* **4**, 565.
- MODIN, S. (1951) *Jernkontorets Ann.* **135**, 169.
- NEHRENBERG, A. E. (1950) *Trans. Metall. Soc. AIME* **188**, 162.
- NICHOLSON, M. E. (1954) *Trans. Metall. Soc. AIME* **200**, 1071.
- PARCEL, R. W. and MEHL, R. F. (1952) *Trans. Metall. Soc. AIME* **194**, 771.
- PETERMANN, J. and HORNBOGEN, E. (1968) *Z. Metallkd.* **59**, 814.
- PULS, M. P. and KIRKALDY, J. S. (1972) *Metall. Trans.* **3A**, 2777.
- RATHENAU, G. W. and BOAS, G. (1954) *Acta Metall.* **2**, 875.
- RIDLEY, N. (1984) *Phase Transformations in Ferrous Alloys*, eds. A. R. Marder and J. I. Goldstein, p. 201, The Metals Society of AIME, Warrendale, PA.
- ROBERTS, G. A. and MEHL, R. F. (1943a) *Trans. Metall. Soc. AIME* **154**, 318; (1943b) *Trans. Am. Soc. Met.* **31**, 613.
- SCHEIL, E. and LANG-WEISE, A. (1937) *Arch. Eisenhuettenwes.* **11**, 93.
- SHPLET, G. J. (1988) *Phase Transformations-87*, ed. G. W. Lorimer, p. 438, The Institute of Metals, London.
- SMITH, G. V. and MEHL, R. F. (1942) *Trans. Metall. Soc. AIME* **150**, 211.
- TURNBULL, D. (1948) *Trans. Metall. Soc. AIME* **175**, 774.
- VERHOEVEN, J. D. and PEARSON, D. D. (1980) *Metall. Trans.* **11A**, 247.
- ZELDOVICH, J. B. (1942) *J. Exp. Theor. Phys. USSR* **12**, 525.

## CHAPTER 18

# *Order–Disorder Transformations*

### 80. THE FORMATION OF A SUPERLATTICE

In most cases, the appearance or disappearance of a superlattice corresponds to a first-order phase change, and a change of temperature from above the critical disordering temperature to below it usually leads to a nucleation and growth reaction. However there are also a sizeable number of ordering reactions which are homogeneous or continuous in the sense that nucleation is not involved. In either case, provided the initial and final structures are both single-phase the ordering process involves atomic rearrangement but no long-range diffusion and, in this respect, the kinetics are similar to those of polymorphic changes. There are sometimes complications caused by the formation of a network of antiphase domains which coarsen slowly in a manner analogous to grain growth.

An order–disorder transformation differs from other transformations inasmuch as kinetic processes within a one-phase region have also to be considered. When the temperature of a pure component or a random solution is altered, the corresponding changes in the lattice parameter, density, etc. are affected immediately. In a superlattice phase, the degree of long-range order is a function of temperature but changes in long-range order require thermally activated processes and hence take a finite time. This means that a change of temperature within an ordered phase may produce a situation which is strictly metastable.

The mechanism by which the degree of order is adjusted to its equilibrium value is not necessarily the same as that utilized in the more extensive change from the disordered to the ordered state. In principle, of course, there are also kinetic effects associated with temperature changes within a disordered phase or within any non-ideal solid solution, as the short-range order (or clustering) depends on the temperature. In most cases the short-range order at high temperatures will attain an equilibrium value very rapidly. Changes in short-range order may also be produced by external mechanical stress, and the finite relaxation time required for the change is utilized in internal friction measurements.

There are marked differences in the rates at which superlattices form from disordered solid solutions in different alloys. Thus the superlattice in  $\beta$ -brass forms in a time too short to be measurable, whilst several hours are required to produce the  $\text{Cu}_3\text{Au}$  superlattice and

more than a week is required for  $\text{Ni}_3\text{Mn}$ . These differences are not attributable entirely to the trivial effect of atomic mobilities. The three examples quoted have roughly equal disordering temperatures and the melting point temperatures are also comparable. It is, however, probable that the different reaction rates are due to the differences in the crystallography of these changes. The  $\beta$ -brass transition is the simplest of structural changes (A2-B2) and this corresponds to a second-order change. The other changes of type A1-L1<sub>0</sub> are necessarily classical first-order changes. The important distinction has already been made, namely that antiphase domains can form a network in the  $\text{Cu}_3\text{Au}$  and  $\text{Ni}_3\text{Mn}$  structures but not in the  $\text{CuZn}$  structure. The time taken to order  $\text{Cu}_3\text{Au}$  at lower temperatures seems to represent the growth of antiphase domains rather than the nucleation of regions of long-range order.

There are essentially two alternative descriptions of the mechanism of an ordering or disordering reaction. In the first theory, the formation of a superlattice is assumed to be a nucleation and growth process, beginning at discrete centres and growing outwards until the disordered regions have been consumed. The second hypothesis treats the reaction as a continuous process of atomic interchange, proceeding simultaneously in all parts of the assembly and being analogous to a homogeneous chemical reaction. Obviously the two mechanisms are related to the thermodynamic nature of the change, a nucleation and growth reaction being more probable with a first-order phase change and a continuous transformation being more probable with a second-order phase change. Continuous changes may be treated by the methods developed originally to deal with spinodal decomposition, i.e. by using concentration waves of very short wavelength.

The theory of nucleation requires some modifications when applied to order-disorder changes. The close relations of the two structures means that extensive rearrangements of atomic positions are not required and the existence of short-range order above the disordering temperature may mean that suitable groups of atoms are available to form embryos. The interfacial energy between the ordered and disordered phases will probably be very small, depending on the change in lattice parameter and symmetry. These factors all tend to make nucleation easier, and to compensate for the rather smaller driving force which is all that is likely to be available.

Because surface and strain energies are likely to be less important than in other transformations, the nuclei will probably form throughout a grain rather than at grain boundaries and dislocations. The orientations of the product regions will be fixed by the parent lattice and, particularly if both structures have cubic symmetry, the product regions forming in one grain will eventually unite to form a single grain of product. As they are independently nucleated, however, the different regions will not necessarily be in phase with each other, and a network of antiphase domains may thus be formed. The antiphase domains will be large just below the temperature at which ordering begins but, at lower temperatures, the reaction will be faster and the antiphase domains smaller.

When there is a change of symmetry on the formation of the superlattice, the situation is a little more complex. In a cubic-tetragonal change, for example, the tetragonal  $c$  axis could be parallel, or nearly parallel, to any one of the three cubic axes, and differently nucleated regions in one cubic grain will meet along antiphase boundaries which represent

differences in orientation. The structures produced in certain alloys which undergo transformations of this type are very similar to the structures produced by some martensitic transformations.

There is ample evidence that the formation of a superlattice may be a slower process just below the transition temperature and especially in a region where ordered and disordered phases coexist. Figure 18.1 shows some data obtained by Rhines and Newkirk (1953) for the change in electrical resistivity of  $\text{Cu}_3\text{Au}$ . At  $389.5^\circ\text{C}$  the change is scarcely perceptible until a few days have elapsed, whereas at  $381^\circ\text{C}$  the reaction is quite rapid. Properties such as electrical resistivity used to follow the kinetics of reactions have to be interpreted with great care; in this case, the resistivity probably depends not only on the relative size of the domains, but also on the degree of long-range order appropriate to the temperatures.

There are many experimental investigations which have established the existence of antiphase domains in ordered structures. The earliest work was probably that of Sykes and his collaborators (Sykes and Evans, 1936; Sykes and Jones, 1936; Jones and Sykes, 1938) on  $\text{Cu}_3\text{Au}$ . Measurements were made of the electrical resistivity and the diffraction line

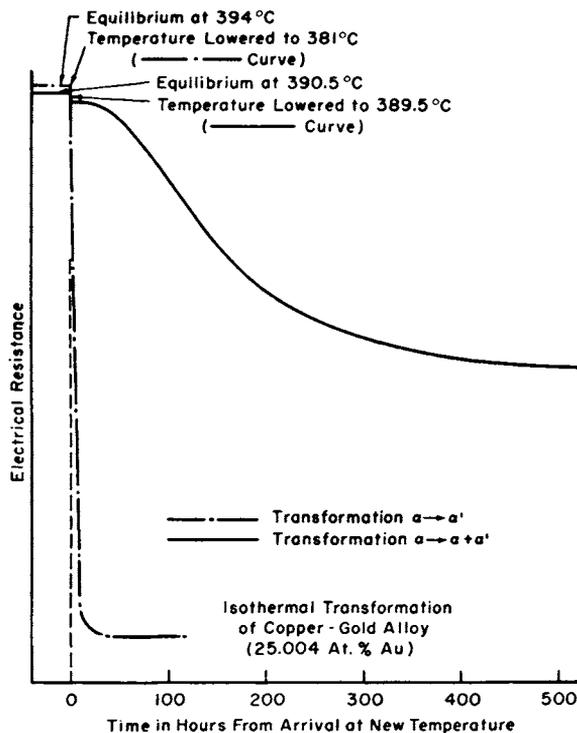


FIG. 18.1. Dependence on temperature of the rate of isothermal transformation of an alloy near  $\text{Cu}_3\text{Au}$  (after Rhines and Newkirk, 1953). The rate of transformation, measured by the electrical resistivity, is much smaller for the  $\alpha \rightarrow \alpha + \alpha'$  transformation at  $389.5^\circ\text{C}$  than it is for the  $\alpha \rightarrow \alpha'$  transformation at  $381^\circ\text{C}$ .

breadths of the superlattice lines as a function of time at a given temperature. The disordered specimens were quenched to room temperature and then heated to 298, 346 or 576°C. After varying times, the specimens were again quenched to room temperature where the resistivity and line breadths were measured. The progress of the ordering reaction, as measured by the decrease in resistivity, was found to be associated with the growth of antiphase domains, measured by the line breadths. Jones and Sykes calculated the domain size by using Scherrer's formula for diffraction from small particles and found that the apparent size varied considerably with the indices of the reflection considered. A more accurate treatment by Wilson (1942) largely resolved the discrepancy and showed that the results were consistent with a hypothesis that gold-gold contacts are avoided at the {100} interfaces. The data of Jones and Sykes indicate that, at the temperatures concerned, the alloy possesses some long-range order after very short annealing times, although the resistance has hardly changed. This means that, at these temperatures, nucleation is rapid and does not constitute a rate-limiting factor. Nearly all the resistance change is associated with the growth of the domains, which are typically about 7 nm in linear dimensions after 10 min at 237°C and grow to over 60 nm after about 12 h. An interesting effect is that a specimen cold-worked after quenching gave a smaller transformation rate, according to both resistivity and line breadth data, than did a standard sample.

Thin film transmission electron microscopy was first used by Fisher and Marcinkowski (1961) to reveal the domain structure of Cu<sub>3</sub>Au, and this work has been repeated by a great many later workers. Fisher and Marcinkowski found that the domains were roughly cubical blocks with boundaries parallel to {100} planes. After 75 h at 380°C, the mean domain size was about 75 nm, in good agreement with the values quoted above. As the temperature is lowered the degree of order within the domains increases.

The mechanism of the growth of the domains may involve highly coordinated atom movements at the interface or possibly the motion of a step across the interface. Such a step is equivalent to a partial dislocation of the superlattice structure or a lattice dislocation of the disordered structure. The domains grow until they are of the order of size of the disordered grains; their growth does not seem to be much impeded by accidental factors such as impurity particles.

A domain size as small as 7 nm in the early stages of reaction suggests that a limit is being approached in which nucleation is so prevalent that a nucleated reaction degenerates into a continuous reaction. Lipson and his colleagues (Lipson, 1950) suggested very early on that a homogeneous reaction was a more natural explanation of the change than a nucleated reaction. Some experiments by Edmunds and Hinde (1952) provided some evidence in favour of a continuous mechanism. The results of the measurement of diffracted X-ray intensity in reciprocal space agreed qualitatively with the predictions of one of Wilson's models, but there were some discrepancies that were attributed to over-simplification in the model used for theoretical analysis. In the early stages of ordering, the results are similar to those of Cowley (1950), who investigated the short-range order above the transition temperature. This similarity is one argument for a continuous mechanism. Further support was provided by experiments of Taylor *et al.* (1951), who studied the

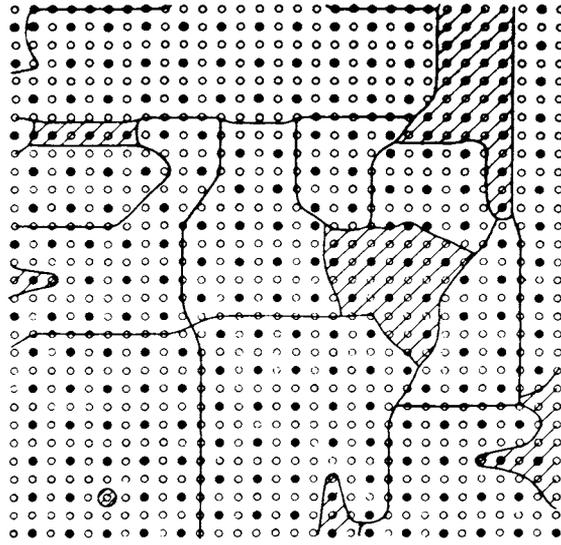


FIG. 18.2. Two-dimensional optical mask used to simulate the X-ray diffraction pattern from a partially ordered  $\text{Cu}_3\text{Au}$  alloy (after Taylor *et al.*, 1951). Open circles represent copper atoms and filled circles gold atoms. The shaded areas cannot be regarded as domains.

optical diffraction patterns from masks in which the copper and gold atoms were represented by holes of different sizes. The problem was made two-dimensional by projecting two adjacent  $\{100\}$  planes onto a single plane. The initial arrangement was random and the order was increased in stages by interchanging pairs of atoms so as to reduce the number of gold-gold contacts. It was found that the resultant optical diffraction patterns were very similar to the X-ray results of Cowley in the early stages and to those of Edmunds and Hinde in the later stages.

Figure 18.2 shows part of one of the masks at an intermediate stage. The development of order in the manner just described has produced a number of antiphase domains and these tend to meet across  $\{100\}$  faces in such a way that gold atoms are not in contact with each other. Thus the formation of the antiphase structure commonly observed is possibly a more or less accidental feature of a continuous ordering process and does not necessarily imply a nucleated reaction. Similar results have been given by Steeple and Lipson (1956) for the alloy  $\text{CdMg}$ . However, the demonstration of geometrical feasibility did not include consideration of the slight change in lattice symmetry.

## 81. ORDERING AS A CONTINUOUS REACTION: CHANGES IN LONG-RANGE ORDER

A quantitative kinetic theory for continuous ordering was first given by Rothstein (1954) and by Dienes (1955) who showed how the variation of the long-range order

parameter with temperature introduced modifications into the kinetics expected of a second-order chemical reaction. These modifications are such that large fluctuations are required under certain conditions and the general behaviour is then similar to that of a nucleated reaction, although the concept of nucleation is not introduced specifically into the theory. Thus, according to the work of Dienes, the observation of sigmoidal transformation curves and other features normally taken to indicate nucleation do not preclude a homogeneous mechanism for ordering, especially in  $A_3B$  alloys. Dienes emphasized that his work only proved a homogeneous reaction to be possible, and it does not prove that nucleation is not really needed. In a later development by Nowick and Weissberg (1958), the theory is restricted to small changes in long-range order within the region of stability of the superlattice.

Dienes' treatment is based on the rather special assumptions of the Bragg-Williams model. The interchange of  $A$  atoms and  $B$  atoms from right sites of the superlattice to wrong sites is treated as a reaction as on p. 220. The rate of change of order with time is then

$$dx_w/dt = -k_f(x_A - x_w)(x_B - x_w) - k_w x_w^2 \quad (81.1)$$

where, as before,  $k_f$  and  $k_w$  are rate constants for the forward and backward reactions respectively. The only interchanges which are physically possible are between neighbouring atoms and, in terms of the reaction rate theory of Chapter 3, these rate constants are

$$\begin{aligned} k_f &= \nu_f(z_B/x_B) \exp -(W + \Delta_a h_{AB})/kT \\ k_w &= \nu_w(z_B/x_B) \exp -(\Delta_a h_{AB})/kT \end{aligned} \quad (81.2)$$

where  $z_B$  is the number of  $B$  sites adjacent to an  $A$  site,  $W$  is the Bragg-Williams energy introduced on p. 220 and  $\Delta_a h_{AB}$  is the activation energy for the interchange of a wrong  $A$   $B$  pair. The frequencies  $\nu_f$  and  $\nu_w$  for the interchange of right and wrong pairs respectively are assumed to include the entropies of activation; the two frequencies may be considered to be equal if the entropies of activation are equal.

At equilibrium,  $dx_w/dt = 0$  and this leads to eqn. (26.9). From eqn. (81.2), it follows that the Bragg-Williams equation for  $L$ , eqn. (26.11), depends essentially on the further assumption that  $\nu_f = \nu_w$ . After expressing  $x_w$  in terms of  $L$ , eqns. (81.1) and (81.2) now give for the rate of change of order

$$dL/dT = k_w [x_A x_B (1-L)^2 - (\nu_f/\nu_w) \{L + x_A x_B (1-L)^2 \exp(-W/kT)\}] \quad (81.3)$$

This is the basic kinetic equation of the homogeneous theory. It differs from an ordinary second-order chemical equation because  $W$  depends on  $L$ . Because of this dependence, the equation is transcendental and may only be solved numerically for assumed values of the parameters.

The procedure used by Dienes is to adopt the Bragg-Williams assumption  $W = W_0 L$  and to evaluate  $dL/dt$  as a function of  $L$  for assumed values  $\nu_f = \nu_w = 1 \times 10^{13}$ ,  $\Delta_a h_{AB}/k = 5000$  K and  $W_0/k = 1000$  K. (Because of a higher entropy in the disordering change,  $\nu_f$  will probably be greater than  $\nu_w$ ; the effect of this was examined separately and shown to be small.) For an  $AB$  alloy, the assumptions correspond to a critical temperature of 250 K and the form of the curves of  $dL/dt$  versus  $L$  is shown in Fig. 18.3;

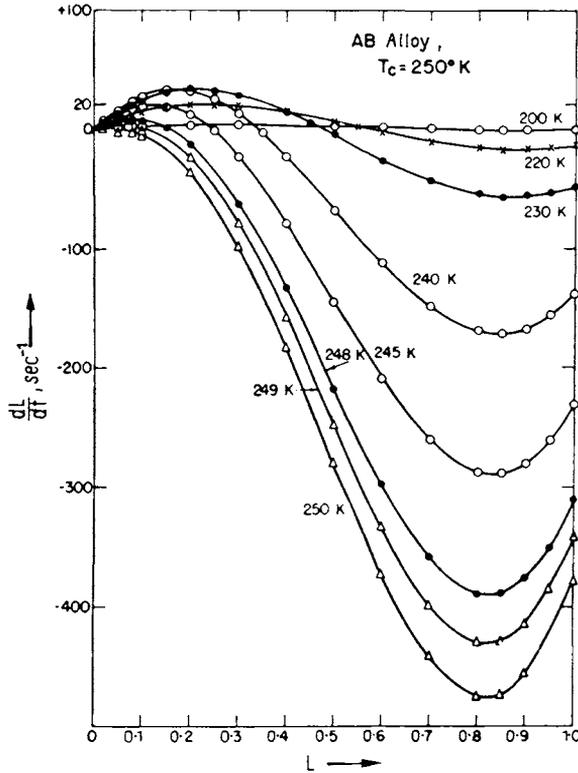


FIG. 18.3. Curves of  $dL/dt$  versus  $L$  for an  $AB$  alloy using arbitrary parameters as described in the text (after Dienes, 1955).

these curves cross the  $dL/dt=0$  axis at values of  $L$  which are the equilibrium values according to the Bragg-Williams theory. The ordering rate at  $L=0$  is always zero but  $dL/dt$  has a positive slope, so that the small fluctuations required to initiate transformation should persist and induction periods are not expected. For any finite amount of long-range order present initially, the ordering rate is also finite at time zero.

In spite of the positive values of  $dL/dt$ , the isothermal curves of  $L$  against time, obtained by numerical integration from an assumed initial amount of order, have sigmoidal shape. The general complex behaviour of these curves is shown in Fig. 18.4, in which it is assumed that the initial order is  $L=0.02$ . The disordering behaviour is rather similar, but the rate of disordering (negative values of  $dL/dt$ ) is finite for any value of  $L$  greater than the equilibrium value. The temperature dependence of the ordering rate is such that there is a maximum rate at some temperature below the transition temperature, as observed experimentally.

Figure 18.5 shows Dienes' calculations for an  $A_3B$  alloy using similar values of the parameters. For quite a large temperature region below the ordering temperature  $T_2$ ,  $dL/dt$  is negative for initial values  $L=0$ . This corresponds to a virtual disordering rate,

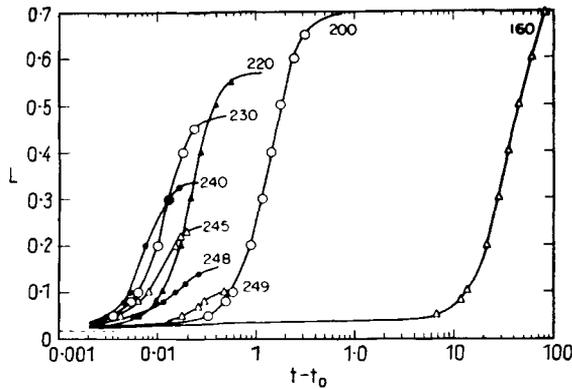


FIG. 18.4. Computed isothermal curves of long-range order versus time for an  $AB$  alloy using arbitrary parameters (after Dienes, 1955).

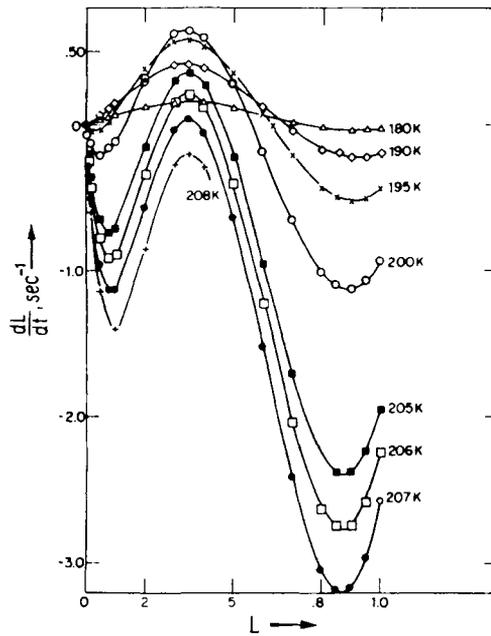


FIG. 18.5. Curves of  $dL/dt$  versus  $L$  for an  $A_3B$  alloy using arbitrary parameters as described in the text (after Dienes, 1955).

which is impossible, and large fluctuations are thus needed to produce a degree of order which will grow. This is akin to a nucleation event, although the theory does not include any arguments based on the size of the region considered. Long induction periods and other characteristics of a nucleation and growth reaction are thus necessary for an  $A_3B$  alloy in a certain temperature range, although this concept has not been deliberately

introduced into the theory. The disadvantages of Dienes' theory are the necessity for numerical solutions and the unrealistic assumption of the Bragg-Williams theory that  $W = W_0L$ . Whilst this may be a reasonable approximation for the change from zero to nearly complete order in a few homogeneous transformations, its more general utility is probably confined to small changes in  $L$ . For this purpose, the theory has been considerably improved by Nowick and Weisberg (1958). These authors show how to solve the fundamental kinetic equation under the restriction that the change in  $L$  from its initial value to its final equilibrium value is small in comparison with unity. Although this means that the theory cannot be applied to changes from the disordered state to the fully ordered state, this may not be a serious limitation in many cases where such changes are nucleated.

Nowick and Weisberg expanded eqn. (81.3) in a Taylor series about the equilibrium value  $L = L_c$ , and showed that the solution can be written in the form

$$\frac{(1-L)/(1-L_c) + \lambda}{1 + \lambda} = \frac{\coth}{\tanh}(\alpha_1 t + \beta_1) \quad (81.4)$$

where  $\alpha_1$ ,  $\beta_1$  and  $\lambda$  are constants. The coth function on the right-hand side of this equation applies when  $L > L_c$ , i.e. the order is decreasing.

Equation (81.4) is similar to an equation

$$(1-L)/(1-L_c) = \coth(\alpha_1 t + \beta') \quad (81.5)$$

which was derived by Rothstein (1954) under rather restrictive conditions (valid at low temperatures) that  $(1-L)$  and  $(1-L_c)$  are both small compared to unity and  $W$  is constant. Nowick and Weisberg showed, however, that eqn. (81.5) approximates closely to eqn. (81.4) under very much wider conditions which are always likely to be reached in practice. The correspondence breaks down only if  $\lambda$  approaches minus one, and this can only result from a much more rapid dependence of ordering energy on the degree of order than has ever been observed.

Rothstein's equation (81.5) is very convenient for experimental investigations of the kinetics of a change in the magnitude of long-range order. For small changes in  $(1-L)$ , the physical property concerned may be assumed to be linearly proportional to  $(1-L)$ , so that

$$P = P_0 + C_{28}(1-L)$$

where  $P$  is the measured value of the property and  $P_0$  is the value for perfect order. Equation (81.5) then becomes

$$(P - P_0)/(P_c - P_0) = \coth(\alpha_1 t + \beta_1) \quad (81.6)$$

and the constant  $C_{28}$  does not affect the kinetics. If eqn. (81.4) is used,  $C_{28}$  must first be determined by simultaneous measurements of  $P$  and  $L$ .

Various investigations have shown that eqn. (81.5) represents changes in long-range order below the critical temperature quite well. The earliest work was that of Burns and Quimby (1955), who measured electrical resistivity. Later work by Weisberg and Quimby (1958) showed that the variation of Young's modulus with time after an abrupt increase or decrease in temperature also fitted this equation. Feder *et al.* (1958) measured changes in

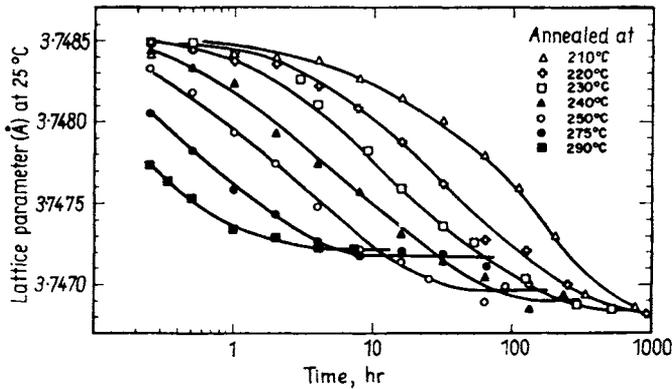


FIG. 18.6. Change of lattice parameter with time, measured at 25°C for  $\text{Cu}_3\text{Au}$  specimens initially quenched from 368°C (after Feder *et al.*, 1958).

lattice spacing at room temperature after stepwise anneals at various temperatures using specimens initially ordered at 308°C. Figure 18.6 shows some of their results; the equilibrium lattice parameter is smaller at lower temperatures (greater order) but is attained more slowly.

In analysing these results and the earlier data, Feder *et al.* first established that all three measured properties were linearly dependent on each other, thus providing some justification for the use of eqn. (81.6). The hyperbolic functions do not fit the data exactly and, in particular, a different value for the constant  $\alpha_1$  is obtained by fitting the coth function to the early stages of the change as opposed to the later stages. According to this theory,  $\alpha_1$  depends only on the temperature and not on the instantaneous order, so that it is a more convenient rate constant than  $dL/dt$ . Nowick and Weissberg developed an approximate expression for  $\alpha_1$  which has the standard form of a frequency factor and an exponential with an activation energy of  $\Delta_s h_{AB} + (1/2)W_0$ . An analysis of results for  $\text{Cu}_3\text{Au}$  showed an activation energy of about 2 eV, which is in agreement with anelasticity measurements of the activation energy for self-diffusion. The frequency factor is about  $5 \times 10^{13}$ , which is of the expected magnitude. Thus the theory was claimed to fit the experimental results for  $\text{Cu}_3\text{Au}$  rather well.

More recent work on ordering (Khachatryan, 1978; de Fontaine, 1975) is based on the idea of concentration waves and especially the notions introduced in the classic works of Landau (1937) and Lifshits (1942). By adopting a generalized Bragg-Williams model (pairwise interactions out to arbitrary coordination shells), the molar free energy  $F_M$  is written as a sum of atomic interactions which is then expanded as a Taylor series to give (to fourth order)

$$F_M = Nf_0 + 1/2 \sum f_0''(\mathbf{p}' - \mathbf{p})\xi(\mathbf{p})\xi(\mathbf{p}') + (1/3!)f_0''' \sum \xi^3(\mathbf{p}) + (1/4!)f_0'''' \sum \xi^4(\mathbf{p}) \quad (81.7)$$

where  $\mathbf{p}$  is a lattice vector,  $f_0 = \bar{\epsilon}_0 c_A c_B + kT(c_A \ln c_A + c_B \ln c_B)$ ,  $f_0'' = 2\bar{\epsilon}(\mathbf{r})$  [ $\mathbf{r} \neq 0$ ],  $f_0''' = -kT(c_A^2 - c_B^2)/c_A^2 c_B^2$ ,  $f_0'''' = 2kT(c_A^3 - c_B^3)/c_A^3 c_B^3$  and  $\bar{\epsilon}(\mathbf{r})$  is the pair interaction potential [see eqn. (23.1)]

$$\bar{\epsilon}(\mathbf{r}) = 1/2[\bar{\epsilon}_{AA}(\mathbf{r}) + \bar{\epsilon}_{BB}(\mathbf{r}) - 2\bar{\epsilon}_{AB}(\mathbf{r})]$$

and

$$\bar{\epsilon}_0 = \sum \bar{\epsilon}(\mathbf{r})$$

Khachaturyan (1983) and de Fontaine (1975) show that it is advantageous to use Fourier transforms on this equation in order to write the free energy in reciprocal space rather than in direct space. The second-order terms are diagonalized in the new representation, and this greatly simplifies the treatment in circumstances where the higher-order terms may be neglected. This applies to high temperatures, where there is mainly disorder, and to small amplitude perturbations of the ordered state. The behaviour of the system is then governed mainly by the Fourier transform of  $\bar{\epsilon}(\mathbf{r})$ , which is written  $\bar{\epsilon}(\mathbf{h})$ , where  $\mathbf{h}$  is a reciprocal lattice vector. In particular, the points at which  $\bar{\epsilon}(\mathbf{h})$  has a minimum value are important as the corresponding wave-vectors minimize the harmonic part of the free energy. Absolute minima of the  $k$  space potential can usually only occur at or near so-called special points which depend on the symmetry properties of the structure. By identifying these special points, various superlattice structures can be predicted (Clapp and Moss, 1966; Khachaturyan, 1978).

Landau and Lifshits listed various symmetry rules which determine whether or not a given transition can be thermodynamically of the second order. When applied in a systematic manner to various known superlattice transitions, it is found that most of them can only be first-order; the best-known exception is the A2-B2 transition in (disordered) b.c.c.  $AB$  alloys. However, although this may indicate that nucleation is required for transformations close to the ordering temperature, it does not necessarily prevent the continuous growth of order at temperatures well removed from the order-disorder temperature. Continuous ordering is now known to occur in many alloy systems. An alloy with long-range order may be regarded as having frozen-in composition waves of very short wavelength (of the order of atomic distances). Continuous ordering, like spinodal decomposition, may form from a disordered solid solution by the gradual increase in amplitude of an appropriate composition wave. Ordering not infrequently occurs in association with phase separation, as in Fe-Al alloys (Allen and Cahn, 1976). The coherent equilibrium diagram for this system is shown in Fig. 18.7. Somewhat unusually, it contains a tricritical point where the second order-disorder transition (disordered B2 structure) intersects the miscibility gap. Figure 18.8 shows schematic free energy versus composition curves which would allow this to happen. An alloy of composition  $c$ , present as a solid solution, is metastable to fluctuations in composition, but if it orders from  $P$  to  $Q$  it may then decompose to give the equilibrium state of which the ordered compound  $A_x B_y$  is one component. The spinodal decomposition cannot take place without the prior ordering and Allen and Cahn called this situation a conditional spinodal. They discuss various other possible transitions on the assumption that the free energy versus

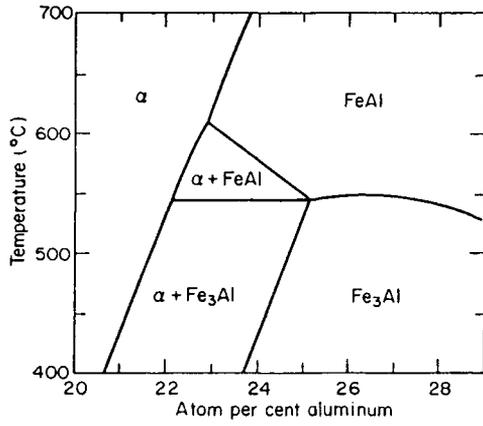


FIG. 18.7. Phase diagram of FeAl system (after Allen and Cahn, 1976).

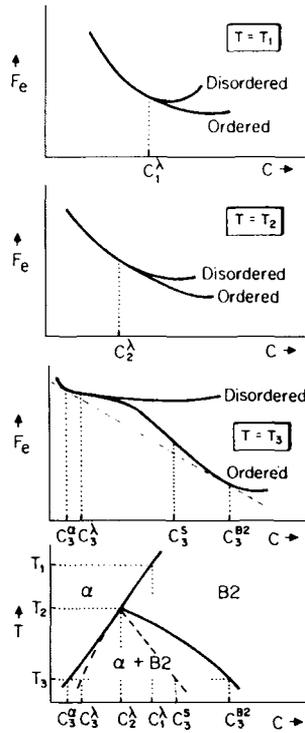


FIG. 18.8. Schematic free energy versus composition curves (after Allen and Cahn, 1976).

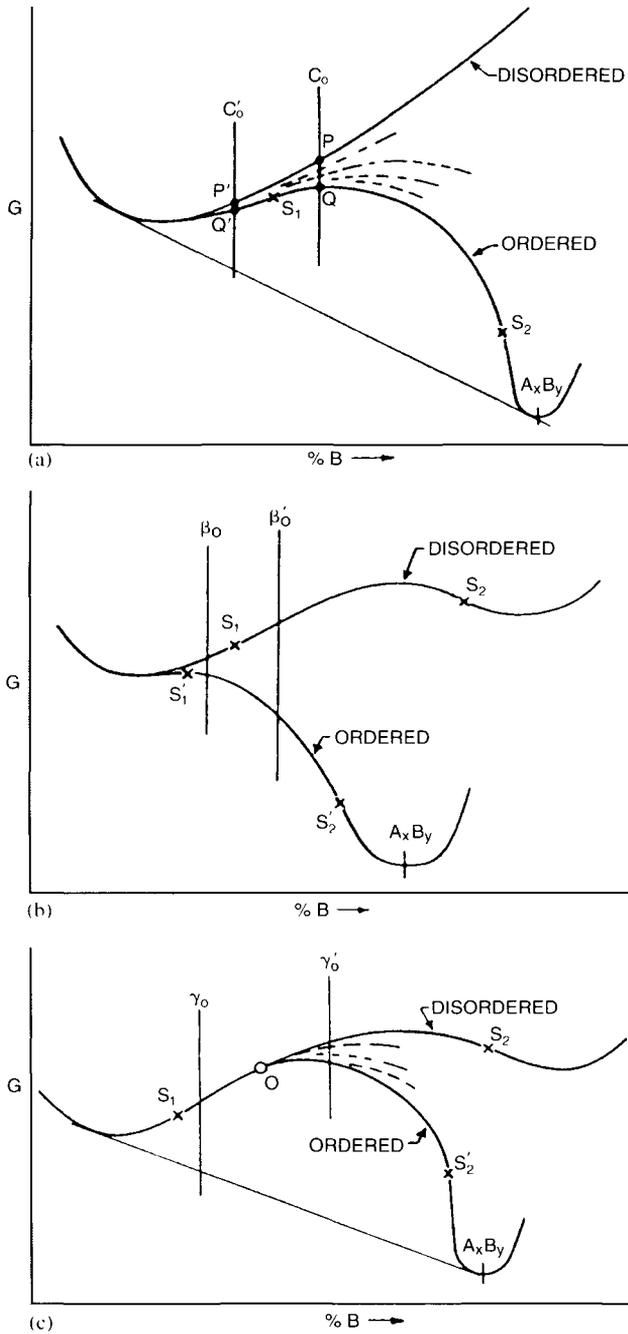


FIG. 18.9. Alternative forms of free energy versus composition curves (after Soffa and Laughlin, 1984).

composition curves are as shown in Fig. 18.8 and also assuming that phase boundaries may be extrapolated into the two-phase regions where they represent metastable equilibrium.

Continuous ordering transformations, often associated with spinodal decomposition, have been located in Cu-Zn, Fe-Al, Fe-Be, Fe-Mn, Fe-Cr, Nb-Zr and Ni-Mn alloys and in various ternary alloys. In a comprehensive review, Soffa and Laughlin (1984) point out that a slight modification of Fig. 18.8 to the forms shown in Fig. 18.9 opens up other possible routes to the final equilibrium structure. In principle, ordering may either precede or follow spinodal decomposition, depending on the details of the free energy curves; however, Allen and Cahn pointed out that ordering will almost always be the faster of the two reactions, so that it will be expected usually to be the first stage in any multi-step process.

#### REFERENCES

- ALLEN, S. and CAHN, J. W. (1976) *Acta Metall.* **24**, 425.  
 BURNS, F. P. and QUIMBY, N. L. (1955) *Phys. Rev.* **97**, 1567.  
 CLAPP, C. and MOSS, S. C. (1966) *Phys. Rev.* **142**, 418.  
 COWLEY, J. M. (1950a) *J. Appl. Phys.* **21**, 24; (1950b) *Phys. Rev.* **77**, 669.  
 DE FONTAINE, D. (1975) *Acta Metall.* **23**, 553.  
 DIENES, G. J. (1955) *Acta Metall.* **3**, 549.  
 EDMUNDS, I. G. and HINDE, R. M. (1952) *Proc. Phys. Soc., London* **65**, 716.  
 FEDER, R., MOONEY, H., and NOWICK, A. S. (1958) *Acta Metall.* **6**, 206.  
 FISHER, R. M. and MARCINKOWSKI, M. J. (1961) *Philos. Mag.* **5**, 1385.  
 JONES, F. W. and SYKES, C. (1938) *Proc. R. Soc. London, Ser. A* **166**, 376.  
 KHACHATURYAN, A. G. (1978) *Prog. Mater. Sci.* **23**; (1983) *Theory of Structural Transformations in Solids*, Wiley, New York.  
 LANDAU, L. D. (1937) *Sov. Phys. J.* **26**, 545.  
 LIFSHITS, E. M. (1942) *J. Phys. (Moscow)* **6**, 61.  
 LIPSON, H. (1950) *Prog. Met. Phys.* **2**, 1.  
 NOWICK, A. S. and WEISBERG, L. R. (1958) *Acta Metall.* **6**, 260.  
 RHINES, E. N. and NEWKIRK, J. B. (1953) *Trans. Am. Soc. Met.* **45**, 1029.  
 ROTHSTEIN, J. (1954) *Phys. Rev.* **94**, 1429A.  
 SOFFA, W. A. and LAUGHLIN, D. E. (1984) *Solid-Solid Phase Transformations*, eds. H. I. Aronson, D. E. Laughlin, R. F. Sekerka and M. C. Wayman, p. 158, The Metals Society of AIME, Warrendale, PA.  
 STEEPLER, H. and LIPSON, H. (1956) *The Mechanism of Phase Transformations in Metals*, p. 77, The Institute of Metals, London.  
 SYKES, C. and EVANS, H. (1936) *J. Inst. Met.* **58**, 255.  
 SYKES, C. and JONES, F. W. (1936) *Proc. R. Soc. London, Ser. A* **157**, 213.  
 TAYLOR, C. A., HINDE, R. M., and LIPSON, H. (1951) *Acta Crystallogr.* **4**, 261.  
 WEISBERG, L. R. and QUIMBY, S. L. (1958) *Phys. Rev.* **110**, 338.  
 WILSON, A. J. C. (1942) *Proc. R. Soc. London, Ser. A* **181**, 360.

## CHAPTER 19

# *Recovery, Recrystallization and Grain Growth*

### 82. ANNEALING OF COLD-WORKED METALS

This chapter is concerned with the changes which take place when a cold-worked metal is heated at a higher temperature. In some respects, these changes are similar to thermodynamic phase changes; processes of nucleation and growth may be involved and the growth stage is characterized by the migration of internal interfaces. As emphasized in Chapter 1, the main difference is in the nature of the driving energy of the reaction.

After deformation at a sufficiently low temperature, a metal or alloy specimen becomes harder and various typical mechanical properties, such as the yield strength, are increased in magnitude. The strain hardening is a result of the increase in the number of dislocations and other defects which hinder the glide of dislocations. Most other physical properties are appreciably changed by the deformation and, when the specimen is subsequently heated, return gradually to values characteristic of well-annealed material. If the original grain structure of the material, as measured by ordinary optical or other techniques, remains undisturbed, the changes in physical and mechanical properties produced by annealing are generally known as recovery. At higher temperatures, most metals will undergo a discontinuous change in grain structure, known as recrystallization. New crystal grains form within the deformed regions and gradually consume the original grains. Primary recrystallization is usually followed by uniform grain growth, or by highly selective grain growth (secondary recrystallization), but the change in most physical properties is virtually complete by the end of primary recrystallization.

Experiments have shown that only a fraction of the energy expended in plastically deforming a metal at low temperatures is stored in the metal, the remainder being dissipated irreversibly as heat. The most reliable data (Clareborough *et al.*, 1955, 1956; Titchener and Bever, 1958; Bever *et al.*, 1973) indicate that between 1% and 15% of the energy may be stored, but the results vary considerably with the metal itself, its purity, and the nature, amount and temperature of the deformation. Maximum reported values of the stored energy for different nominally pure metals given large plastic strains at room temperature range from  $90 \text{ J g atom}^{-1}$  for gold to almost  $900 \text{ J g atom}^{-1}$  for nickel. The energy usually increases with increasing impurity and with decreasing temperature, and is also probably greater in polycrystalline materials than it is in single crystals given comparable strains.

The stored energy represents the energy of the various defects left in the material after deformation, and its existence shows that the deformed state is thermodynamically metastable. The driving force for the recovery and recrystallization processes is the extra free energy of the cold-worked state and this cannot readily be measured. However, a major part of the stored energy will be attributable to the introduction of extra dislocations into the system, and the entropy changes are then negligible, so that the free energy will not differ appreciably from the internal energy or enthalpy. At low temperatures, point defects may make an appreciable contribution to the stored energy and the associated entropy is not then negligible but, even under these conditions, the change in free energy should be considerably greater than half the change in enthalpy.

The driving force for recovery and recrystallization is smaller than the energies utilized in most phase transformations but larger than the energies stored in the grain boundaries of a polycrystalline material.

Assuming a boundary free energy of  $500 \text{ mJ m}^{-2}$ , the boundary energy is about  $1 \text{ J g atom}^{-1}$  for a grain diameter of  $10^{-5} \text{ m}$  and is thus only about 1% of the stored energy of deformation. The two energies would only be comparable for grains of nanometre size. It follows that, in a heavily worked material, the migration of boundaries during recrystallization will be governed entirely by the stored strain energy unless this has already been released by recovery processes. During grain growth or secondary recrystallization, however, the interfacial free energy of the boundaries is the sole driving force for the reaction.

The study of recovery processes is of great importance in determining the type and concentration of structural defects present in the material, and the way in which they interact. If there were a few simple types of non-interacting defects, each with a specific activation energy for motion and hence removal, the form of isochronal or isothermal recovery curves would be relatively simple. In practice, recovery experiments are often very difficult to interpret unambiguously, especially when several types of defect have been introduced by low temperature deformation or by radiation damage. This has led to considerable emphasis on simpler experiments such as the recovery of the excess resistivity due to quenched-in vacancies where only one or two types of defect are present in appreciable concentrations. Some recovery processes occur at very low temperatures; in material irradiated with neutrons at 4.2 K, some recovery (probably involving the recombination of close interstitial–vacancy pairs) takes place at 30–40 K.

Recovery processes in which individual defects are removed or trapped do not come within the general definition of transformations (see Section 1) and are not considered further. During an isothermal reaction of this kind, the rate of reaction continually decreases and the reaction is homogeneous throughout the specimen. However, the development of more refined X-ray and metallographic techniques led to the discovery of structural changes taking place before recrystallization. These changes, generally described as polygonization, involve the breaking up of the original grains into a number of smaller subgrains of almost identical orientation, leaving the main grain structure undisturbed. It is now clear that a considerable fraction of the recovery of mechanical properties prior

to recrystallization must be attributed to polygonization, which may be so effective that true recrystallization is not observed.

Observations on the formation of substructures, first by optical metallography and later by X-ray methods, have a long history, but the nature of the structure was not understood until the development of the dislocation theory of small angle boundaries. The name polygonization comes from the pioneering experiments of Cahn (1949) who bent single crystals of zinc about an axis normal to the slip direction and subsequently annealed them. The crystal surface changed from a uniform curve to a series of sharply tilted regions, with a sub-boundary normal to the slip plane at each tilt. The process is illustrated in Fig. 19.1; the bending introduces a high density of edge dislocations of one sign into the crystal and, on annealing, these dislocations lower their energy by assembling into walls which become simple tilt boundaries. Polygonization no longer has this simple geometrical significance but is used for any process in which the dislocations are assembled into stable arrays.

Cahn's experiment suggests that subgrains will form in any deformed specimen if the deformation is inhomogeneous in the sense that local regions are bent even though no macroscopic bending moment is applied. Such local bending is to be expected whenever there is more than one operative glide system, and it is possible to produce substructures in most cubic metals after simple tensile deformation. Conversely, single crystals of hexagonal metals like zinc or cadmium can be given large deformations in single glide and will then recover their original properties without either recrystallization or polygonization.

Crussard (1944) first observed polygonization in aluminium by X-ray methods when stretched crystals were heated to 350–400 °C. He termed this "recrystallization in situ", and further observed that the subgrains increased in size and became less numerous on subsequent annealing at higher temperatures; this subgrain growth is accompanied by a decrease in yield stress. Later, more refined X-ray techniques were introduced by Guinier and Tennevin (1948) and by Lambot *et al.* (1953) and it became evident that finer substructures were present at an earlier stage of annealing and even immediately after deformation at room temperature. These conclusions have been amply confirmed by X-ray microbeam experiments (Kellar *et al.*, 1950) and by transmission electron microscopy (Hirsch *et al.*, 1957) and it is clear that, for this metal, polygonized structures are formed during the deformation. In severely deformed metal, little or no subgrain growth is observed until temperatures of the order of 200 °C are attained.

In contrast to the behaviour of aluminium, there is little or no evidence of a regular polygonized structure in deformed copper, nickel, iron or stainless steel (Gay and Kelly, 1953; Gay *et al.*, 1954; Whelan *et al.*, 1957). The results of electron microscopic examination of thin foils of deformed metals were briefly summarized on pp. 313–316. They show that a distinction must be made between f.c.c. materials like stainless steel, brass or copper–aluminium alloys, all of which have very low stacking fault energies, and f.c.c. metals like copper, silver and gold of intermediate stacking fault energy. In alloys of low fault energy, the dislocations are confined to slip planes for small deformations, and pile up against the grain boundaries. After larger deformations, complex networks

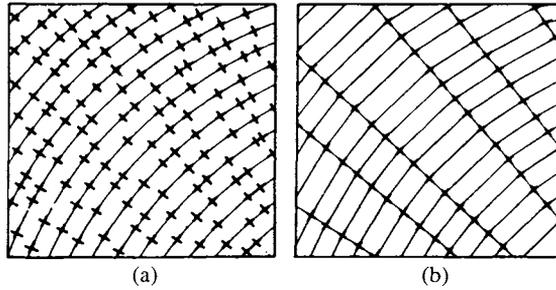


FIG. 19.1. Dislocation rearrangement during polygonization (after Cahn, 1950): (a) as bent; (b) after annealing. The crosses represent excess positive dislocations.

of dislocations begin to form and it becomes more difficult to recognize the slip planes. In the case of copper and similar metals, rather irregular networks begin to form in the early part of the deformation, and these build up into a cell structure which consists of regions relatively free from dislocation surrounded by dense dislocation walls or tangles. The individual cells are misorientated from one another, but they contain more dislocations than the rather perfect structures produced in aluminium after large deformation, and they do not have sharp boundaries. Experiments on silver and copper showed that a "clean" subgrain structure is also produced in these metals by deformation at higher temperatures. Bailey and Hirsch (1960) measured the stored energy of cold work in silver and found it to be equal to the self energy of the dislocations in the dense tangles, thus indicating that any stress fields from the pile-ups are relaxed. The mean cell sizes reported by Warrington (1961) for copper increase from  $0.6\mu\text{m}$  at  $-196^\circ\text{C}$  to  $0.9\mu\text{m}$  at  $0^\circ\text{C}$  and  $1.8\mu\text{m}$  at  $400^\circ\text{C}$ , and the corresponding thicknesses of the boundary regions are  $150\text{nm}$  at  $-196^\circ\text{C}$  and  $0^\circ\text{C}$ , and  $50\text{nm}$  at  $400^\circ\text{C}$ .

There is good evidence that the differences in the structures of various deformed f.c.c. metals and alloys correlate reasonably well with their respective stacking fault energies. In aluminium, with a high energy, the activation energy for cross-slip is low, and dislocations are frequently seen in the electron microscope to cross-slip. The process of edge dislocation climb, possibly using vacancies produced by the deformation, is also easier when the dislocations are not extended, and this may contribute to the formation of sharp dislocation walls.

Dislocations in metals of intermediate fault energy are not able to cross-slip until the effective local shear stress reaches a rather high value characteristic of each metal. The observed cell structures show that some dislocations must leave their original slip planes in this way, and that stress concentrations caused by incipient pile-ups are at once relaxed by the secondary slip systems. Easy cross-slip appears to be essential if a polygonized structure is to be produced by deformation alone.

The formation of a substructure was studied by Jacquet (1954) using etch pit methods. In his work on  $\alpha$ -brass, he noted that dislocations delineated the slip lines after light deformation, but formed along sub-boundaries and also in the interior of subgrains after

annealing. Raising the annealing temperature produced larger and more perfect subgrains with fewer dislocations in the interior.

Many experiments (see, e.g., Wood, 1951, 1953; Thornton and Cahn, 1961) have shown that application of stress during the annealing accelerates considerably the rate of recovery and of subgrain growth; the total amount of recovery is also enhanced. Recovery can also occur during hot working or in creep at high temperatures, in which case it is called dynamic recovery. This is again most pronounced in metals like aluminium and iron in which cross-slip is relatively easy.

The driving force for the formation of a polygonized structure is presumably the strain energy of the dislocations trapped in the material by the deformation. These dislocations may reduce their energy either by mutual annihilation of opposite dislocations (a true recovery process) or by assembly of dislocations into more or less regular arrays which are the boundaries of subgrains. In as much as the slip dislocations usually have a non-random distribution, polygonization involving edge dislocations is only feasible if dislocations climb out of their glide planes. When cross-slip occurs, screw dislocations are probably able to achieve a relatively low energy structure by pure glide movements, and this possibly accounts for the structure observed in aluminium deformed at room temperature. The activation energy for deformation of a polygonized structure should then presumably correspond to that for dislocation climb, except in the presence of an applied stress when it might correspond to that of thermally activated cross-slip. Once a reasonably perfect polygonized structure has formed, most of the dislocations are present in the cell boundaries and the remaining stored strain energy may be regarded as the interfacial energy of these boundaries. The driving force for subgrain growth is then just this surface free energy, and the situation is analogous to ordinary grain growth except that the interfaces have been introduced by deformation. This emphasizes that the distinctions made between surface and strain energy for convenience are often somewhat arbitrary.

No detailed model for the process of subgrain growth appears to have been suggested. In considering the motion of a small angle grain boundary, it is necessary to distinguish the case of the symmetrical low angle tilt boundary from a more general boundary. The symmetrical tilt boundary is the simplest example of a glissile interface (Section 38) which can move under mechanical or thermal stress without the need for thermal activation. This prediction of dislocation theory has been directly verified by Washburn and Parker (1952) and by Li *et al.* (1953) using boundaries induced in single crystals of zinc and observed on a cleavage surface. They found that boundaries moved under applied stress in the direction predicted, and that two parallel boundaries can unite to form a single boundary. This happens because if the two boundaries have different tilt angles, the boundary of smaller angle is more mobile and may overtake the other. On reversing the stress, the composite boundary splits into two boundaries again if the original boundaries represent tilts of the same sign, but not if they have opposite tilts. In the latter case, the opposite dislocations presumably annihilate and splitting into two boundaries would require creation of new dislocations. In the former case, it is not easy to understand why the boundary should split again unless the two boundaries did not fully coalesce. The experimental results for the variation of boundary mobility with misorientation across the boundary are shown in

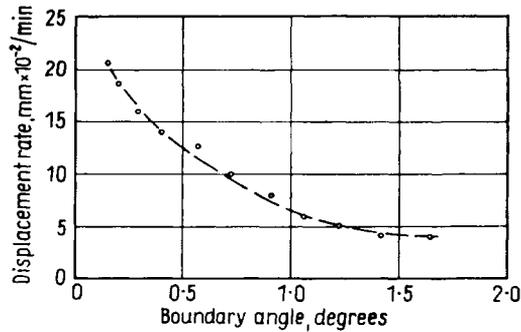


FIG. 19.2. The displacement rate of edge dislocation boundaries in zinc as a function of boundary angle under the action of a constant shear stress of 9.19 p.s.i. at 350 C (after Bainbridge *et al.*, 1954).

Fig. 19.2; correspondingly the minimum shear stress to move the boundary increases with increasing misorientation.

An observation of considerable interest is that, at room temperature and below, an increasing shear stress is required to maintain the motion of the boundary and the tilt across the boundary decreases as it moves (Bainbridge *et al.*, 1954). This suggests that the boundary encounters obstacles which hinder its motion and which are gradually swept up as the boundary migrates. Under these circumstances, the motion of the boundary will be jerky. At higher temperatures, the tilt angle remains constant, and the boundary moves continuously under a constant applied stress. Presumably, dislocations or other obstacles are then being overcome with the help of thermal energy. The rate at which obstacles are overcome will determine the mobility of the boundary, and the rate of motion, being governed by a thermally activated process, will be sensitive to the temperature.

The single tilt boundary is glissile in principle, because it contains a single set of parallel dislocations of the same Burgers vector which can glide together as a unit. Most low angle boundaries consist of intersecting dislocations of more than one Burgers vector; motion by glide of all these dislocations must involve climb, and this requires processes with activation energy about equal to that for lattice diffusion. Most low angle boundaries have limited mobilities which increase as the misorientation increases, in contrast to the simple tilt boundary discussed above. The marked effect of applied stress on subgrain growth (it has little or no effect on normal grain growth) is a measure of the extent to which processes like stress-induced boundary motion and dislocation cross-slip can contribute to subgrain growth. Hibbard and Dunne (1957) have given a detailed survey of the mechanisms of polygonization.

### 83. PRIMARY RECRYSTALLIZATION

The reorganization of a deformed matrix to form a new structure of roughly equi-axed grains is generally described as primary recrystallization. Under most circumstances,

the process corresponds to the formation of new strain-free grains which grow into the distorted matrix until the latter completely disappears. The stored energy of cold work is removed in proportion to the volume of the material which has recrystallized, and the new grains are very much softer, as may be verified by microhardness measurements. At constant temperature, any dimension of a growing grain is observed to increase linearly with the time; if this straight line relationship is extrapolated backwards to zero grain size, it has a positive intercept on the time axis. For this reason, recrystallization is usually treated as a nucleation and growth reaction. As discussed on p. 17, an operational definition of the nucleation rate can be obtained by measuring the growth of a large number of individual grains and then finding how many per unit volume originate in any fixed small time interval.

It follows from the above discussion that this idealized view of recrystallization cannot always be correct. Calorimetric measurements generally show that a considerable fraction of the stored energy is released in recovery processes prior to primary recrystallization. Thus part of the above difference between copper and nickel is attributable to the high mobility in point defects in copper at room temperature. For deformation at room temperature, the proportion of the stored energy varies from 0% for high purity copper to 70% for commercial purity copper. Some of the original stored energy is necessary for recrystallization. Some earlier results indicated that the whole of the stored energy can be released before appreciable softening begins, but they have not been confirmed. In some cases, it does seem possible to remove virtually all the stored energy before recrystallization, but the material then will not recrystallize.

Various experimental investigations of the kinetics of recrystallization have been made and the results are usually expressed in the form of  $\zeta$   $t$  curves, where  $\zeta$  is the volume fraction of the material which has recrystallized. In accordance with the nucleation and growth description, the reaction curves are sigmoidal, the reaction rate first increasing and then decreasing with time. The results may readily be described in terms of the standard Avrami equation but, in the early investigations, it became evident that there are some discrepancies between theory and experiment. The value of  $n$  [see eqn. (4.11)] is between one and two but seems to vary with the material and the external constraints. Such a low value indicates one- or two-dimensional growth with no effective nucleation (see Table IX) (p. 546), but new grains are not observed to form either as rods or platelets. Vandermeer and Gordon (1959, 1963) studied the recrystallization of zone-refined aluminium after deformation by 40%. They found  $n = 2$  and new grains were nucleated in clusters at particular sites along grain boundaries; these nuclei were apparently present at the beginning of the anneal and started growing immediately with no induction period. Long and narrow new grains grew into the matrix at a steady rate. In contrast to this work, however, Anderson and Mehl (1945), using less pure aluminium and a smaller deformation, concluded that the nucleation rate varied with time after an initial induction period and that growth of new grains was isotropic.

According to Vandermeer and Gordon, the Avrami equation is not a good approximation for alloys or impure metals. They found systematic deviations in aluminium to which 0.04% copper had been added. The retardation was attributed to

recovery as the authors showed that recovery and recrystallization overlap. With a lengthy recovery anneal at a lower temperature, the overlap was eliminated, and recrystallization was then retarded. The deviation in an alloy increases with the solute content and at fixed composition is greater at lower temperatures. Recrystallization has a higher activation energy than recovery, and this energy is apparently increased more by addition of copper to the aluminium than is the recovery energy of activation.

The work of Stanley and Mehl (1942) and Anderson and Mehl (1945) was intended partly to separate the nucleation and growth contributions to the overall kinetics. The growth rate was found to be linear and, by making the questionable assumption that this was the same for all grains, it was possible to deduce the nucleation rate from the statistical distribution of grain diameters at different times. The experimental result obtained from this work was that  $I$  increases sharply with time according to an equation of the form

$$I = C_{20} \exp(ht)$$

where the  $\exp(ht)$  term is larger than  $C_{20}$  for all observable times. This result is difficult to reconcile with those of Vandermeer and Gordon, and it is also not consistent with either steady-state nucleation theory or initial nucleation site saturation. Various other workers have measured only the overall kinetics. Decker and Harker (1950) and Seymour and Harker (1950) used X-rays and described their data in terms of the Avrami equation with  $n$  changing from three to two as the temperature was reduced. The results suggested that all nuclei were present at the beginning of recrystallization growth, being three-dimensional for  $n=3$  and two-dimensional for  $n=2$  (see Table IX) (p. 546). Obviously there are other possibilities with non-zero nucleation rates and growth rates decreasing with time, but these seem inherently less probable. The main importance of several investigations seems to be in confirming that recrystallization may be formally described as a nucleation and growth process.

Much attention has been focused on the mechanism of nucleation. It is at once evident that although classical homogeneous nucleation is possible in principle, in practice the probability of this kind of nucleation is zero. For a spherical nucleus, the critical nucleus size would be

$$r_c = 2\sigma v/g_s \quad (83.1)$$

where  $g_s$  is the strain energy per atom. The interfacial free energy is of the order of  $500 \text{ mJ m}^{-2}$  for a high angle boundary, and  $g_s/v$  is less than  $45 \text{ J g atom}^{-1}$  for most metals, giving  $r_c > 0.1 \mu\text{m}$ . This is improbably large for a region produced by thermal fluctuations and, as the corresponding free energy for formation of such a nucleus is  $\Delta G_c = 4\pi r_c^2 \sigma/3$ , the rate of nucleation will be negligible. This is the most favourable situation; detailed measurements of dislocation density by thin film electron microscopy gave a critical radius of  $0.5 \mu\text{m}$  for silver deformed 25% in tension (Bailey, 1960) and of  $1 \mu\text{m}$  for a silicon-iron crystal rolled to 70% reduction in area (Hu and Szirmai, 1961). The energy stored in a single crystal (and hence the radius) is very dependent on orientation. In a similarly deformed but differently orientated single crystal, crystallization nuclei stopped growing at diameters of about 40 nm, presumably because they had entered regions of the matrix

where the mean driving force was so low that they were below the critical size for growth (Hu, 1959).

This last result shows that the difficulty cannot be avoided by postulating local regions in which the stored energy per unit volume is much larger than its average value. If such regions exist, they can only produce nuclei which continue to grow if the stored energy falls off with distance very gradually, so that the actual radius of the growing grain always exceeds the critical radius of eqn. (83.1). In principle, the energy is a maximum near a dislocation core, but this could not lead to a nucleus of greater than atomic dimensions. In the regions of high dislocation density (cell walls) produced by severe deformation, the critical nucleus size may be reduced by a factor of five (Bailey and Hirsch, 1960) but it could not grow out of the cell walls unless it attained a size appropriate to growth in the cell regions of much lower dislocation density and driving force.

Because the failure of the classical theory of homogeneous nucleation for recrystallization phenomena arises from the small driving force per unit volume, no modification which is based on heterogeneities or defects smaller than the critical nucleus size above will confer any advantage. Thus theories of heterogeneous nucleation by random thermal fluctuations are equally inapplicable, and the only acceptable theories are those which postulate the prior existence or ready formation of very large nuclei (or very small grains) which can grow into the matrix. The most highly developed theory of this type is due to Cahn (1950) following an earlier suggestion by Burgers (1947).

In Cahn's theory, strain-free regions of small size form by polygonization in the early stages of annealing. This is not a nucleation process in the true sense, but arises from thermally activated rearrangement as already discussed. The apparent incubation period obtained by extrapolation of experimental curves like that of Fig. 1.2 is not the time to form a nucleus, but is indicative of an initial period of slow growth. Thus the theory is equally applicable to the growth of small, nearly stress-free cells formed directly during deformation if these exist.

This type of theory illustrates the warning on p. 17 that reaction curves like Fig. 1.1 or growth curves like Fig. 1.2 do not necessarily indicate true nucleation and growth transformations. In a sense, recrystallization is now only a growth process, although the operational definition of the nucleation rate remains valid. The reasons for the increase of growth velocity with size were further elaborated by Cottrell (1953). The initial boundary between a growing region and its surroundings will be a low angle type and, in accordance with the discussion on p. 836, will be relatively immobile except in the special circumstance where a glissile tilt boundary is formed. As the boundary moves outwards, it collects up individual dislocations and dislocation tangles (cell walls) and the orientation of the surrounding matrix deviates increasingly from that of the recrystallized region. When the misorientation becomes large enough, a high angle boundary of relatively large mobility has been formed. This theory emphasizes the importance of the variables giving boundary migration rates, not only for grain growth in the later stages of recrystallization but also throughout the recrystallization process. Intensive studies of these variables have been made, but evidence has also accumulated that the process of recrystallization does not begin in quite the way envisaged by Cahn. Hu (1962, 1963)

suggested a related model in which neighbouring subgrains coalesce by the removal (by glide and climb) of the dislocations forming the boundary between them. In this way a subgrain much larger than its neighbours would form and would be able to grow into the surrounding matrix.

Electron microscopic and other observations (see, e.g., Bollman, 1958–9) sometimes show recrystallization to begin before distinct subgrains have formed and to develop independently of polygonization. Observations on thin foils of silver (Bailey, 1960) showed directly that recrystallization begins from high angle boundaries present in the initial structure before deformation. This is consistent with a large body of data showing that it is much easier to recrystallize polycrystalline specimens than single crystals.

Bailey's mechanism for recrystallization, also observed by Beck and Sperry (1950) on a much coarser scale, is a form of strain-induced boundary migration. He considered a simple model in which there is a local difference in dislocation density on the two sides of a grain boundary. This causes it to bulge out into a spherical cap, creating a dislocation-free region. If a circular region of boundary of radius  $R$  is involved in this initial bulging, the criterion for growth to be able to occur is then [by analogy with eqn. (83.1)]

$$R > 2\sigma v/g_s \quad (83.2)$$

giving, for silver deformed 20% in tension,  $R > 0.5\mu\text{m}$  as above. Further evidence that recrystallization in other f.c.c. metals is dependent on grain boundary migration after small to medium deformations is given by Bailey and Hirsch (1962). Figures 19.3–19.6 show unrecrystallized regions and recrystallized regions in both copper and nickel. In these experiments recrystallization begins below the temperature at which dislocations can move by climb, and polygonized structures are not formed.

Bailey and Hirsch considered a specific model in which a growing grain bulges out between fixed points on a boundary (Fig. 19.7). If the volume of the grain is  $V$  and its surface area is  $O$ , the rate of transfer of atoms across the boundary is

$$(O/v^{2/3})(v/kT)[g_s - \sigma v(dO/dV)] \exp(-\Delta_a g^*/kT)$$

where  $v$  is the volume per atom and  $O/v^{2/3}$  is taken as the number of sites at which atoms can cross the interface. This rate of transfer is equal to  $(1/v)(dV/dt)$  so that

$$dV/dt = O v^{1/3} f [g_s - \sigma v(dO/dV)] \quad (83.3)$$

where  $f = (v/kT) \exp(-\Delta_a g^*/kT)$ . Expressing  $O$  and  $V$  in terms of the fixed length  $R$  and the angle  $\alpha$ , eqn. (83.3) becomes

$$d\alpha/dt = (v^{1/3} f/L) p [g_s - (2\sigma v/R) \sin \alpha] (1 + \cos \alpha) \quad (83.4)$$

The condition for growth to occur is that  $d\alpha/dt > 0$  for all values of  $\alpha$ , which gives eqn. (83.2) above.

By integrating eqn. (83.4) a rather complex expression for the time  $t$  taken to reach a particular growth condition (specified by a parameter  $x = \tan(1/2)\alpha$ ) is obtained. This growth equation is most conveniently expressed in terms of a growth parameter  $\beta = 2\sigma v/Rg_s$ . The volume of a growing grain is proportional to  $(x^2 + 3x)/\beta^3$  and Bailey

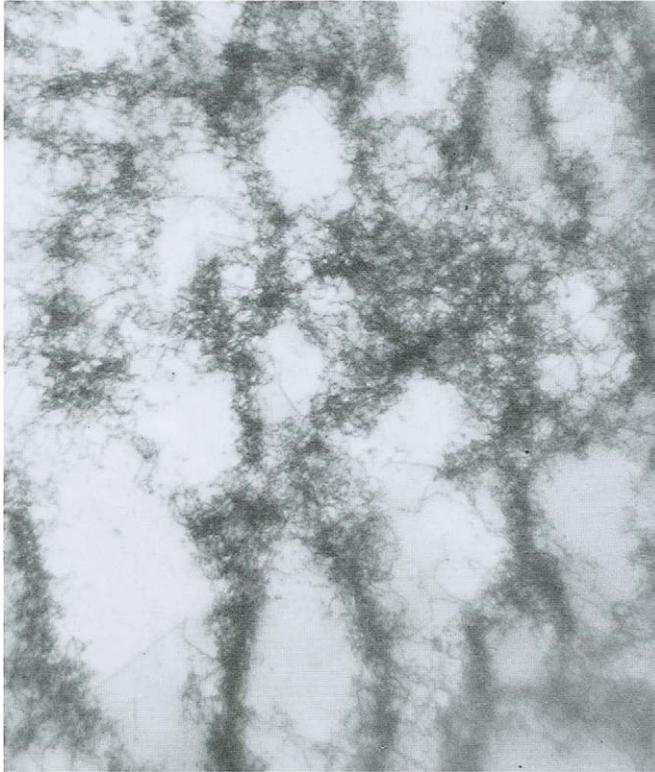


FIG. 19.3. Unrecrystallized region within one grain of polycrystalline copper deformed 16% in tension and annealed for 5 min at 234 C (original magnification  $\times 30,000$ ; courtesy of Bailey and Hirsch, 1962).

and Hirsch investigated the growth laws by plotting this volume against  $t(f/g_s^2/\sigma v^{2/3})$  for different values of  $\beta$ . The slope of such a curve has a minimum value at  $x = 1$  ( $\alpha = \pi/2$ ), so that the rate of increase of a linear dimension in a growing grain is a minimum when the grain is hemispherical. This can give an apparent incubation period, as in growth curves of the type shown in Fig. 1.2. A detailed comparison of the predicted kinetics with the Avrami equation (4.1) leads to the conclusion that various apparent values of  $n$  are possible but that  $n = 3$  for  $x > 1$ .

This theory emphasizes recrystallization purely as a growth process and the temperature dependence is describable in terms of a specific activation energy, that for grain boundary migration. Good agreement is obtained with experimental parameters for the kinetics of recrystallization in silver and copper, and the theory seems more convincing than earlier descriptions, although the assumption of isotropic spherical growth may be unrealistic.

Although there is experimental evidence that the dislocation density may be low in regions where a boundary begins to bulge as in Figs. 19.4 and 19.6, the existence of a such



FIG. 19.4. Grain boundary migration in polycrystalline copper deformed 14% in tension and annealed for 5 min at 234 C (original magnification  $\times 15,000$ ; courtesy of Bailey and Hirsch, 1962).

a large boundary region with maximum driving force across it seems a little unlikely. An alternative suggestion made by Davies *et al.* (1961) depends on the formation of “ledges” on grain boundaries in the interior of the specimen as a result of slip, a ledge being the internal equivalent of a surface slip step.

The ledge theory of recrystallization is illustrated in Fig. 19.8. The slip ledge  $AB$  of height  $d$  tends to straighten at the temperature at which grain boundary migration can occur, and so reduce the total energy by decreasing the grain boundary area. By migrating to  $AC$ , a region of crystal free from dislocations is produced and very low angle boundaries are left behind near  $A$  and  $C$  to accommodate the slight differences in orientation between the old region of the grain  $X$  and the new region which forms the nucleus  $X'$ . This nucleus can now grow forward into the grain provided that  $AC$  is greater than  $2r_c$ . This modification to Bailey's theory uses some of the initial grain boundary energy to provide a driving force for the grain when it is too small to grow under the influence of the average strain energy alone.

It is clear that there will be a critical height of the ledge below which growth is impossible. More detailed consideration, which must include the effect of growth on the

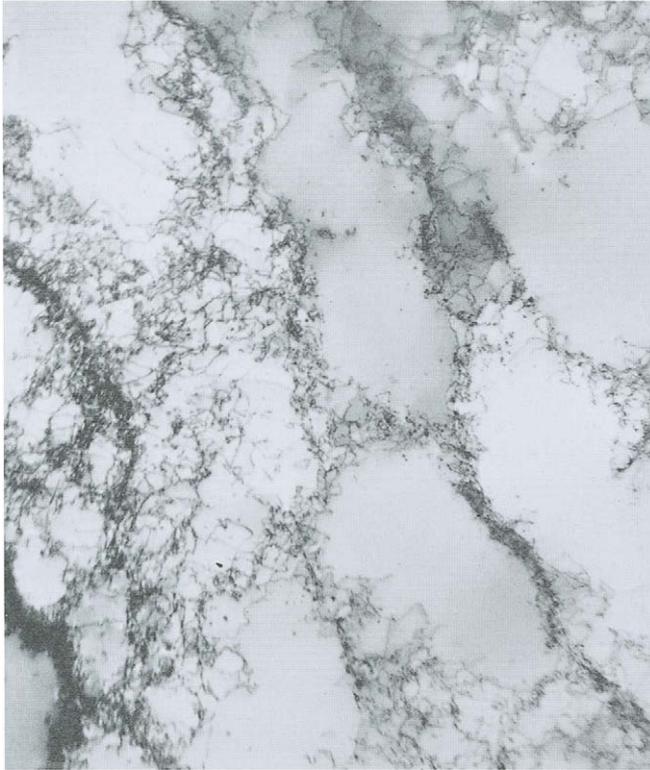


FIG. 19.5. Unrecrystallized region within one grain of polycrystalline nickel annealed for 10 min at 425 C (original magnification  $\times 30,000$ ; courtesy of Bailey and Hirsch, 1962).

low angle boundaries delineating the original high angle boundary, shows that the critical condition is

$$d > 0.1r_c \quad (83.5)$$

This gives a ledge height of about 50 nm for Bailey's data on silver, and this would not have been detected experimentally. Moreover, combination of this theory with Bailey's assumption of a local variation in dislocation density could lead to a still smaller ledge height.

It is still not entirely clear how effective grain boundaries are in promoting recrystallization, but it seems probable that growth begins from them in all moderately deformed polycrystalline metals. In single crystals or very coarse grained material, growth may begin from polygonized areas as postulated by Cahn and much more severe deformation is needed to promote recrystallization under comparable conditions.

Experiments on the growth of stress-free grains into a deformed matrix are relevant also to the growth of such grains in secondary recrystallization or uniform grain growth. The velocity  $\Upsilon$  is expected to depend on the relative orientations of the grains which meet at

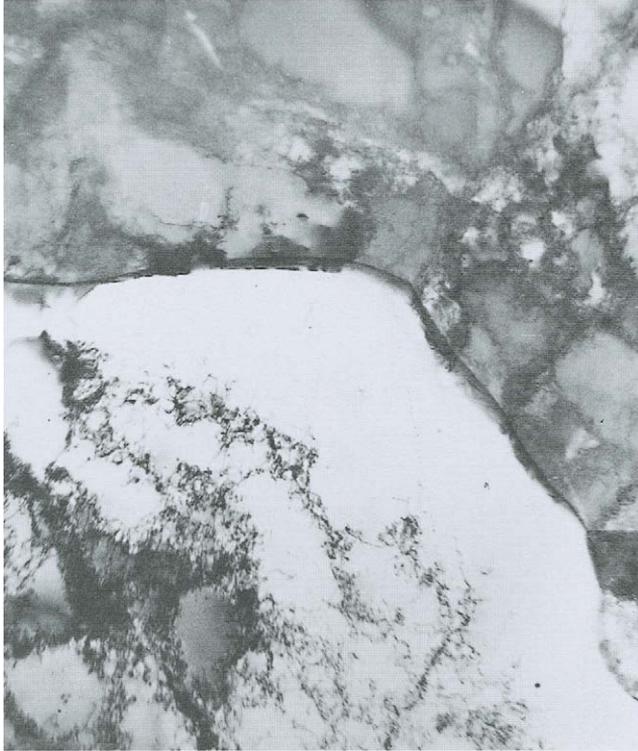


FIG. 19.6. Grain boundary migration in polycrystalline nickel deformed 16% in tension and annealed for 10 min at 425 C (original magnification  $\times 15,000$ ; courtesy of Bailey and Hirsch, 1962).

some interface. This misorientation is best specified as an axis and a minimum rotation about this axis which will bring the two crystal lattices into coincidence. In addition to the three parameters of this relation, the velocity should also depend on two further parameters giving the orientation of the boundary plane.

Two orientation effects may be distinguished. The first is the difference in mobility of large and small angle boundaries already noted. Apart from glissile interfaces, low angle boundaries should migrate much less readily than high angle boundaries because the structures are semi-coherent across the boundary. This prediction has been verified experimentally in various ways. Tiedma *et al.* (1949) and Lacombe and Bergehezan (1949) noted, for example, that small grains remaining inside large recrystallized grains of aluminium all had orientations which were either closely related or twin-related to the enclosing lattice. A fully coherent ( $K_1$ ) twin interface, although glissile under mechanical stress, is apparently very immobile when motion is due to individual atoms crossing the interface. Observations with the thermionic emission microscope (Rathenau, 1952) show that interfaces other than the  $K_1$  interface of annealing twins migrate readily during grain growth in a fully annealed nickel-iron alloy, whilst the fully coherent  $K_1$  interfaces are

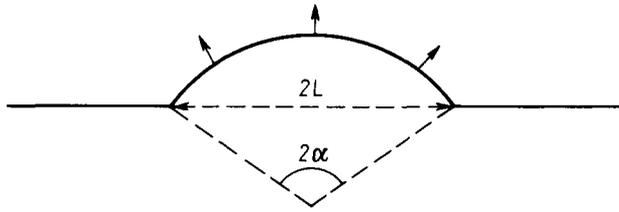


FIG. 19.7. Model for grain boundary migration (after Bailey and Hirsch, 1962).

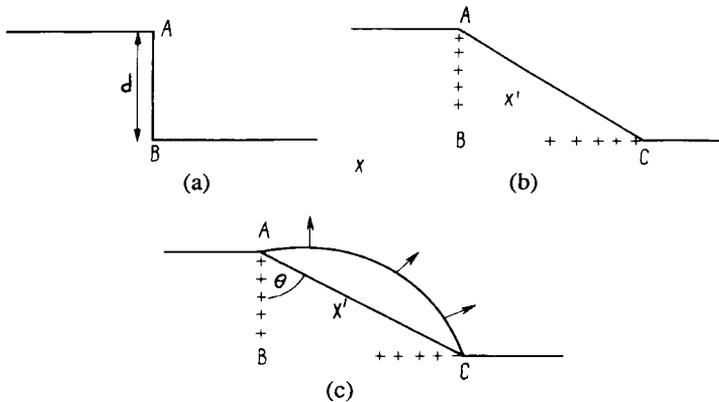


FIG. 19.8. Diagram illustrating the ledge model for the initiation of boundary migration (after Davies *et al.*, 1961). The ledge  $AB$  in (a) moves so as to decrease the grain boundary area, thereby producing a dislocation-free region  $X'$  in (b). The boundary  $AC$  can then migrate into the neighbouring grain as shown in (c). The crosses indicate dislocations remaining in low angle boundaries.

virtually immobile. The very slow coarsening of the subgrain structure during polygonization may also be interpreted as showing that the subgrain boundaries have low mobility, although this is not conclusive evidence because the driving force is also very small.

The second type of orientation effect is important in the theory of recrystallization textures. Experimental results show that not all high angle boundaries are equivalent, and that some special boundaries migrate much more rapidly than do the majority of random high angle boundaries. Beck *et al.* (1950) found that, of the many recrystallized grains forming around a scratch in a lightly deformed single crystal of aluminium, those grains with orientations differing from the matrix by a rotation of about  $40^\circ$  about a  $\langle 111 \rangle$  axis grew much more rapidly than the others. Kronberg and Wilson (1949), studying secondary recrystallization in copper, reported that rapidly growing grains are close to two orientations obtained by rotations of either  $38^\circ$  or  $22^\circ$  about a  $\langle 111 \rangle$  axis. In later work, it became clear that these two orientations, together with a rotation of about  $28^\circ$  about a  $\langle 100 \rangle$  axis, constitute special boundaries which can move rapidly in f.c.c. metals in some

circumstances. These relations are among those which give the highest density of "coincident sites" in the two lattices, as discussed by Kronberg and Wilson (1949) (see pp. 345–349).

The marked effect of impurities on boundary mobility was mentioned above. Very detailed investigations of the dependence of boundary mobility on both purity and misorientation were made by Aust and Rutter (1959a,b) and by Rutter and Aust (1960) using direct measurements of boundary velocities in bicrystal specimens. The bicrystals were obtained by growing one crystal from the melt with a striation substructure (see p. 653) and introducing the second crystal by localized deformation. This second crystal was often the most successful of several which had nucleated so that rapidly growing orientations were obtained when conditions favoured them. It was also found that when coincidence site operations had a higher mobility, they were often introduced into the specimen by a twinning process; this is discussed on p. 349. The crystal grew into the undeformed single crystal matrix under the influence of the energy of the striation substructure, which thus provided a (very small) driving force for measurements of growth velocity. The main series of experiments were made with zone-refined lead, with additions of very small concentrations of tin, silver or gold. The results showed that all solutes markedly decreased the rate of boundary migration, the effects of silver and gold being greater than that of tin.

Special boundaries of high mobility were not observed in pure lead at 300 °C, or in the alloys with silver or gold, but they appeared in alloys with tin contents in the range of 5–40 at. ppm. In later work, it was found that coincidence site boundaries of high mobility form in zone-refined lead at 175 °C; this may be due to the residual effects of some impurity present in very small concentrations but this cannot be stated with certainty. Figure 19.9 shows that when special boundaries exist their velocities decrease less rapidly with solute content than do those of the random boundaries. Later work includes experiments on zone-refined aluminium (Rath and Hu, 1972), again with small driving forces, and on iron (Leslie *et al.*, 1963). Aust (1974) has reviewed his experiments and discussed the effects of both impurities and of slight deviations from the coincidence site orientations. The fact that the velocity is so sensitive to impurities and to orientations in these experiments does not guarantee that the same effects hold in recrystallization where the driving force is very much larger.

If these results can be generalized to other metals and solutes, they imply that special boundaries are able to migrate more rapidly than general boundaries only in the presence of particular solutes in critical concentration ranges. There have been conflicting reports on whether or not similar effects are seen in recrystallization. Liebmann *et al.* (1956) and Liebmann and Lücke (1956) found a pronounced growth selectivity for commercial purity aluminium with the most rapidly moving grains separating regions with relative rotations of 40° about (111), whilst a parallel investigation (Graham and Cahn, 1956) gave no evidence of growth selectivity. Green *et al.* (1959) suggested that the preferred growth was suppressed by dissolved iron. Further results by Frois and Dimitrov (1961) found preferred orientations for coincidence site boundaries in zone-refined aluminium with 2–50 ppm of copper, and Parthasarathi and Beck (1961) reported selective growth in

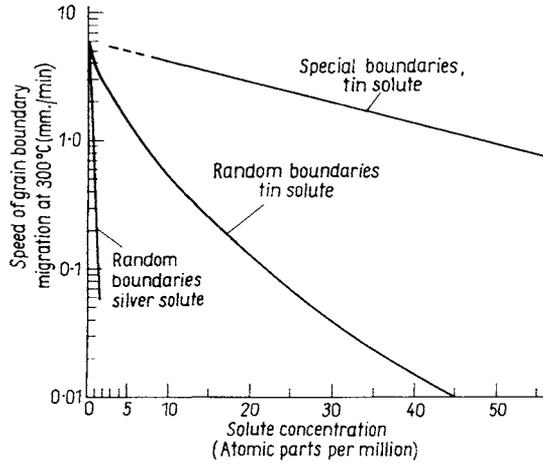


FIG. 19.9. The effects of tin and silver on the speed of grain boundary migration in zone-refined lead (after Aust and Rutter, 1960a).

both high purity aluminium and in an alloy containing iron, silicon and zinc. These specimens were rolled 80% and the authors suggest that, at such high driving forces, the orientation dependence is not nearly so sensitive to solute impurities. One inference from the results of Aust and Rutter and Frois and Dimitrov would seem to be that the conditions for formation of sharp annealing textures are rather stringent, whereas such textures seem to form under most conditions even with appreciable impurity contents.

The results obtained by Aust and Rutter can be rationalized on the basis that solute impurities segregate to grain boundaries and are then dragged along with the boundary under certain conditions. The special boundaries represent cases of rather better fit according to the Kronberg–Wilson model and the interaction energy between a solute atom and the boundary should correspondingly be reduced. Figure 19.9 shows that the interaction of silver with a boundary is much more significant than that of tin. This suggests that there is segregation of tin to the random boundaries only, whereas the higher interaction energy of silver results in segregation to all boundaries. For any given solute, the selective segregation will only be effective over a certain composition range after which there will be segregation to all boundaries.

The most plausible model for growth at a general grain boundary is one in which individual atoms surmount an energy barrier in passing from one crystal to the other, so that the theory of Section 53 should be applicable. The growth velocity is then given by eqns. (53.6) or (53.8), with the driving force for recrystallization or grain growth replacing the chemical free energy difference between the  $\alpha$  and  $\beta$  phases. As already stated, it is plausible to regard the activation enthalpy in this model as appreciably smaller than the experimental activation enthalpy for lattice diffusion. There is good evidence that this is so for lead and it seems well established that in very pure lead and in lead–tin alloys special boundaries move with an activation energy of about  $25 \text{ kJ g atom}^{-1}$  which is independent

of tin content. In contrast to this, the measured activation enthalpies in lead–silver alloys increase with increasing alloy content, reaching  $125 \text{ kJ g atom}^{-1}$  at 0.8 ppm of silver.

The high activation enthalpies are typical of most results obtained from the temperature dependence of grain boundary migration in primary recrystallization, secondary recrystallization and grain growth. The activation enthalpy and corresponding activation entropy both have values appreciably higher than those predicted from the theory of Section 53, and the enthalpy may be up to twice the measured activation enthalpy for lattice diffusion. Mott (1948) suggested a model in which a small group of  $n$  atoms become disordered on one side of the interface and re-order on the other side. The theory was related to his concept of a high angle grain boundary having islands of alternate fit and misfit between the crystals, and the disordering process was compared with the local melting of a small group of atoms. The enthalpy of activation should thus be of the order of  $n\Delta h^{\text{sl}}$  and the entropy of activation of the order of  $n\Delta s^{\text{sl}}$ . Although this accounts for the high enthalpy and entropy, the model seems a little unnatural, and there is no indication of what determines the magnitude of the parameter  $n$ . Moreover, the high activation energy is not related in any way to the impurity content of the material, although later results show that there is a strong correlation with this factor. Aust and Rutter (1959b) discussed this theory and found that  $n$  had to be given values ranging from four to 36 for different lead–tin alloys, and the theoretical velocities were then  $10^2$ – $10^4$  times higher than the experimental velocities. Similar conclusions were reached by Holmes and Winegard (1959), who reviewed results on normal grain growth.

Soon after the discrepancy between measured and expected activation enthalpies was first realized, Turnbull (1951) showed that the effect could be attributed to a retardation of the velocity of the interface by inclusions present in the matrix. For any one state of dispersion of the inclusions, the temperature dependence of the growth rate should give the correct activation enthalpy for the single atom transfer across the interface. However, if the state of dispersion is itself a function of temperature, the total variation with temperature of the velocity could be markedly changed, thus giving an apparent activation energy which is the result of two simultaneously operating temperature-dependent factors.

A quantitative theory of boundary migration in the presence of impurities was developed by Lücke and Detert (1957), who pointed out that impurity atoms segregated to a boundary could cause a drag on that boundary and could well limit its mobility. They assumed that, under normal conditions, the impurity atmosphere has to be carried along with the boundary, the velocity of which is reduced by the necessity for diffusing these atoms behind the boundary. An expression for the growth velocity is obtained in the form:

$$Y = (g_s/kT)(D/\mu) \exp(-W_s/kT) \quad (83.6)$$

where  $D$  is the lattice diffusion coefficient and  $W_s$  the interaction energy between the solute atom and the boundary. An elastic model is used for  $W_s$  and, as this is much smaller than the activation energy for lattice diffusion, the experimental activation energy for growth should be equal to the latter. This is not in agreement with the experimental results. The

theory predicts that at low solute contents or high temperatures, the interface can no longer be held by the impurities and breakaway occurs. The breakaway condition is rather stringent; for example, in lead–tin alloys, Aust and Rutter estimate that, at 300°C, the concentration of tin must be less than 0.4 ppm ( $10^{-5}$  at.%).

As would be expected from the failure to predict the observed activation enthalpy the Lücke–Detert model does not give good agreement with experimental determinations of interface velocities. The discrepancy is about one order of magnitude for lead–tin alloys and two to three orders of magnitude for lead–silver alloys. A much more detailed theory of the impurity drag on a moving boundary has been given by Cahn (1962); this shows that the effectiveness of a particular impurity in slowing down a boundary depends on its diffusion rate relative to the boundary velocity, and its interaction with the boundary. The faster diffusing impurities are more effective when the boundary velocity is large and vice versa. Changes in interaction with the boundary as the impurity content is increased, or as the temperature is lowered, may lead to transitions in boundary behaviour and to high apparent activation energies with jerky boundary motion. Special boundaries are expected to be more mobile than random high angle boundaries, the effect being most pronounced at high boundary velocities.

Recrystallized metals often have a pronounced texture which is different from any deformation texture of the cold-worked material before recrystallization. There is a voluminous literature on this subject, and no attempt will be made to give an adequate summary here. Early work is summarized by Burke and Turnbull (1952) and by Beck (1954). The f.c.c. metals develop recrystallization textures which are related to the prior deformation textures by rotations about  $\langle 111 \rangle$  axes, the angle of rotation often being in the region of  $40^\circ$ . This is the expected result if special boundaries of  $40^\circ$  rotation (see Table VII) have higher mobilities than other high angle boundaries. An orientation relation often observed in recrystallized silicon iron is a  $30^\circ$  rotation about a  $\langle 110 \rangle$  axis which represents a coincidence site lattice relation in the b.c.c. case.

There are two main theories of recrystallization textures. The older “orientated nucleus” theory (Burgers and Tiedma, 1953) supposes that the texture originates, at least in part, because nuclei in certain orientations form more rapidly than others. The “orientated growth” theory (Beck, 1953) assumes that nuclei of all types may form initially, but the texture results from the dependence of the growth rate on the misorientation of the two lattices across an interface. Doherty (1982) expresses the two alternatives as a frequency or a size advantage. Orientated nucleation requires that the frequency of nucleation of the special orientation should greatly exceed the frequency expected from random nucleation. For orientated growth, it is required that grains with the special orientation are larger than other grains, at any stage of the process. Doherty defines the frequency and size by coefficients  $\alpha$  and  $\beta$  respectively which give the ratio of the observed particular orientation to the average and states that if either  $\alpha$  or  $\beta$  is very much larger than unity, this demonstrates that the appropriate condition holds. For many years, the evidence for the rotations quoted above seemed strong. Modern experimental techniques enable  $\alpha$  and  $\beta$  to be readily determined from the microstructure. For many years, the results quoted above seemed to favour orientated growth, especially in deformed single crystals, but more

recently there have been experiments on heavily deformed polycrystalline copper, and high and commercial purity aluminium, which indicate a dominance in the production of cube texture of orientated nucleation with  $\alpha \gg 1$ .

In many experiments on f.c.c. metals, it has been found that only four of the eight equivalent components of a texture corresponding to  $40^\circ$  rotations about  $\langle 111 \rangle$  are actually present in appreciable amounts. Following a suggestion by Burgers, Parthasarathi and Beck (1961) have shown that this can be due to a dependence of the migration rate on the orientation of the boundary itself. When two grains with a common  $\langle 111 \rangle$  direction have an interface normal to this direction (a twist boundary) the interface migrates more slowly along the common direction than do interfaces which contain the common direction. This emphasizes the importance of the local boundary structure, which cannot be characterized by a single parameter such as the reciprocal density of coincidence sites.

#### 84. GRAIN GROWTH, SECONDARY RECRYSTALLIZATION AND TWIN FORMATION

The uniform coarsening of a stress-free material at a high temperature is described as grain growth. Grain growth is normally regarded as a process following primary recrystallization, but in principle it also applies to cast metals. The driving force is the intergranular free energy of the boundaries remaining in the specimen, and an outline description was included in Section 35. The driving force is proportional to  $(1/r_1) + (1/r_2)$ , where  $r_1$  and  $r_2$  are the two principal radii of curvature in mutually perpendicular directions of a local section of boundary. Observations on a planar section may be made by quenching the specimen at intervals during growth or by observing continuously with a hot stage microscope or a field emission microscope. Such observations show that boundaries generally migrate towards their centres of curvature, as required by the above assumption about the driving force. Occasional migration in the opposite direction may be observed, probably resulting from an opposing curvature in a section at right-angles to that under observation.

Although observations of grain growth are normally made on surface sections, it should be noted that this procedure may give unrepresentative behaviour. The free surface itself hinders the motion of a boundary which intersects it, and there is a tendency to form a groove to satisfy the conditions for local free energy equilibrium. Kinetic data from surface observations must therefore be interpreted with caution.

The geometry of grain growth was considered on pp. 327–332. The requirements of local surface free energy at three grain junctions can be fairly readily met by relatively small atomic adjustments so that, under most conditions, the angular relations between three interfaces meeting along an edge will be maintained close to an equilibrium configuration. In order to fulfil this condition most boundaries cannot be planar, and hence will migrate towards their centres of curvature (on average) by a process of atom transfer across the boundaries. This migration will generally be continuous but occasionally unstable situations will arise, leading to rapid readjustments and

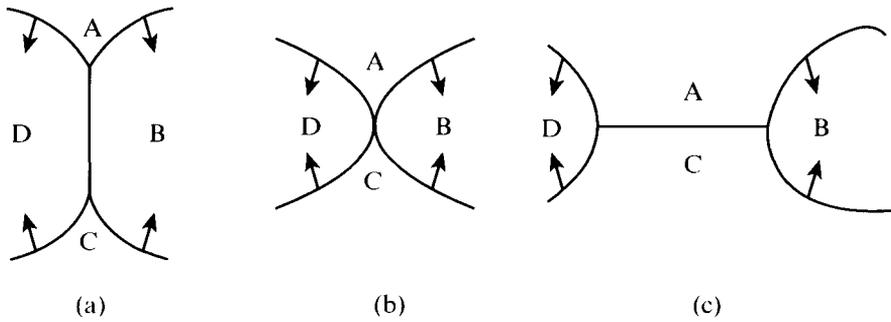


FIG. 19.10. The migration of the grain boundaries in (a) produces the unstable four-grain junction of (b) which immediately splits into two new three-grain junctions (c).

discontinuities in the motion. The three-dimensional equivalent of Fig. 8.1 provides such an unstable situation in which four grains meet along an edge. Figure 19.10 shows how migration of the boundaries towards their centres of curvature gives a four-grain junction which immediately separates into two new three-grain junctions. The growing grains *A* and *C* each acquire a new surface as a result of this process and the shrinking grains *B* and *D* each lose a surface. By means of processes like this, grains occasionally increase or decrease the numbers of their bounding interfaces, and grains with only three interfaces are periodically eliminated completely.

To obtain a model for grain growth, suppose that the driving force in unit volume is proportional to the grain boundary energy and hence to the reciprocal of the mean grain diameter  $L^B$  [see eqn. (35.4)]. If the velocity of boundary migration is given by eqn. (53.6), and substituting  $dL^B/dt = Y$

$$dL^B/dt = (\delta^B v/k)(3.35\sigma v/L^B T) \exp(-\Delta_a g^*/kT)$$

Integrating this equation from the initial grain size  $L^B = L_0^B$  at  $t=0$  gives

$$(L^B)^2 - (L_0^B)^2 = [(3.35\sigma v \delta^B v/kT) \exp(-\Delta_a g^*/kT)]t \quad (84.1)$$

and if  $L_0^B$  is negligible in comparison with  $L^B$ , this simplifies to

$$L^B = Ct^{1/2} \quad (84.2)$$

Experimentally it is found that

$$L^B = Ct^n \quad (84.3)$$

where values of  $n$  ranging from 0.056 to 0.60 have been reported (Burke, 1949; Beck *et al.*, 1948). The failure to follow the theoretical relation (84.2) can be attributed to the effects of impurities or inclusions and, when these are absent, good agreement with eqn. (84.2) is observed (see, e.g., Feltham, 1957; Bolling and Winegard, 1958a; Holmes and Winegard, 1959). In experiments on high purity nickel and zone-refined lead and tin, the measured

activation enthalpies are approximately equal to those for grain boundary diffusion and are much lower than those obtained from commercially pure materials.

Holmes and Winegard (1959–60) investigated the effects of small additions of lead, bismuth, silver and antimony on the grain growth of zone-refined tin, and the results may be regarded as being complementary to those of Aust and Rutter discussed above. Grain growth was retarded to an increasing extent from lead to bismuth to silver as solute, the measured activation energy increasing from 25 kJ g atom<sup>-1</sup> for zone-refined tin to 92 kJ g atom<sup>-1</sup> for an addition of 0.02% bismuth. The growth rates reached a common value near the melting point of tin and Holmes and Winegard proposed modifying the above growth equations by including an additional term  $\exp(-Q/k)\{(1/T) - (1/T^{\text{sl}})\}$ , where  $Q$  depends on the impurity and its concentration and  $T^{\text{sl}}$  is the melting point of tin. The exponent in eqn. (84.3) is quite close to a half in all three of the solutes mentioned. Quite different results were obtained with antimony as solute; the rate of grain growth was markedly reduced close to the melting point of tin and was almost independent of temperature. The results are not believed to support the Lücke–Detert theory but may be in general agreement with a description in terms of the adsorption of solute atoms at the advancing edge of a growing grain (Oriani, 1959).

In principle, grain growth should continue until the whole specimen is a single crystal or at least until the grains have dimensions comparable with the smallest external dimension of the specimen, the grain boundaries then being planar and stable. This is sometimes observed, but more usually a limiting grain size is obtained, after which grain growth virtually ceases. The blocking of grain boundary migration may be achieved by inclusions and direct evidence for this is available. A treatment of the interaction due to a spherical inclusion of Zener and quoted by Smith (1948) is as follows. The force exerted by the inclusion on the boundary resolved normal to the boundary is  $\pi r \sigma \sin 2\theta$  and this has a maximum value of  $\pi r \sigma$ .

If there are  $N$  particles per unit volume, their volume fraction is  $f = 4\pi r^3 N/3$ . A boundary of unit area will intersect all particles within a volume of  $2r$ , i.e.  $2Nr$  particles. Hence, eliminating  $N$ , the number of particles intersecting unit area of a boundary will be  $3f/2\pi r^2$ . Now suppose the boundary is migrating under the curvature driving force  $2\sigma/R$ , where  $R = L^{\text{B}}$  the grain diameter. This force is thus  $2\sigma/L^{\text{B}}$  and the grain will cease to grow when this is just balanced by the force from the inclusions, which is  $3f/2\pi r^2 \times \pi r \sigma$ , giving  $L^{\text{B}} = 4r/3f$ . Thus to a sufficient approximation, the limiting grain diameter is given by the average diameter of the inclusions divided by their volume fraction. Using this assumption, Burke (1949) derived a modified form of eqn. (84.1) as

$$\{(L_0^{\text{B}} - L^{\text{B}})/L_m^{\text{B}}\} + \ln\{(L_m^{\text{B}} - L_0^{\text{B}})/(L_m^{\text{B}} - L^{\text{B}})\} = [C/(L_m^{\text{B}})^2]t, \quad (84.4)$$

where  $L_m^{\text{B}}$  is the limiting grain diameter. In this way, the deviation of the time exponent in eqn. (84.3) from the value of a half may be explained.

A matrix which has undergone primary recrystallization and attained a stable situation may sometimes undergo highly selective growth of one or two grains. This process is secondary recrystallization and is so called because it displays apparent nucleation and growth characteristics in the same way as primary recrystallization. Secondary

recrystallization occurs only when a fine-grained structure with a fairly well-developed texture has been produced by primary recrystallization, and also appears only in material in sheet form. It is essentially a grain growth process and can occur only if small grains of the required orientation are present at the end of primary recrystallization. In f.c.c. metals, the texture produced by secondary recrystallization is related to the primary texture in the same way as the grains of that texture are related to the cold-worked matrix. This does not give a return to the cold-worked texture as all variants of the relation are used, but the observation suggests a growth selectivity theory of secondary recrystallization. The best-known secondary recrystallization texture is the cube texture produced in silicon-iron, and this has been extensively investigated by Walter and Dunn (1959).

Various suggestions have been made about the driving force for secondary recrystallization. They include the presence of stored strain energy in the grains formed by primary recrystallization, a larger surface free energy driving force for certain grains caused by an initial advantage in size, or a higher driving force for the growth of certain grains because of the lower energies of their free surfaces. In addition there is the growth selectivity theory just mentioned in which the driving force is the same for all grains but the presence of one or two boundaries of high mobility in a matrix containing mainly boundaries of low mobility will lead to highly selective growth. The surface free energy hypothesis seems to fit the observation that the secondary recrystallization is found only in sheet specimens, and a direct proof of it was given by Walter (1959) for silicon-iron. A sample was cut at  $10^\circ$  to the plane of a recrystallized sheet and given a heat treatment to promote secondary recrystallization. This developed a texture in which all the cube faces were within  $5^\circ$  of the new plane of the sample as expected from the surface theory. All the other explanations would have required the texture to have the usual relation to the plane of the recrystallized sheet.

It now appears that secondary recrystallization occurs because of a surface free energy driving force, but the texture which appears is modified by growth considerations. This is illustrated by "tertiary recrystallization", which was discovered by Walter and Dunn (1959). In this phenomenon, a secondary recrystallized matrix with a strong texture undergoes further selective growth to give a new texture. In the case of silicon-iron, the secondary cube texture is replaced by a "cube on edge" or  $(110)[001]$  texture, the driving force for which is believed to be the lower surface energy of b.c.c.  $\{110\}$  planes as compared to  $\{100\}$  planes. The tertiary phase is thought not to form directly from the primary recrystallized structure because the grains with the correct orientation for this texture have a low mobility in the primary matrix. There is thus an intermediate stage with the production of the secondary texture by grains which have relatively high mobility in the primary matrix. Both secondary and tertiary recrystallization are essentially processes which refer to changes in texture and can only occur when the appropriate grains are present as a minor component. It is necessary that a rather stable structure should have been achieved either because of the thickness of the sheet or the perfection of the existing structures.

A prominent feature of many f.c.c. metals and alloys is the large number of annealing twins which appear in the recrystallized microstructure. These usually appear

as parallel-sided bands bounded by coherent  $\{111\}$  planes but occasionally the twins terminate within a grain on some other interface. During annealing, the  $\{111\}$  twin interfaces are very immobile, but the other interfaces often migrate and small included twins can disappear in this way.

As described in Section 16, a f.c.c. twin is equivalent to a growth fault and may be considered to arise after a growth accident. If stacking faults exist in the deformed matrix, there is obviously a favourable situation for twin formation when a growing grain reaches a fault. This is supported by a good correlation between the density of stacking faults as measured by X-rays and the frequency of annealing twins after recrystallization. However, growth accidents of other kinds could lead to twins and it seems certain that what is significant is the condition for a minute twin or a monolayer fault to grow into a macroscopic crystal. The standard theory of this is due to Fullman and Fisher (1951) and is illustrated in Fig. 19.11. The twin  $T$  is formed at the three-grain junction where grain  $S$  is growing at the expense of  $U$  and  $V$ . The introduction of  $T$  will decrease the total interfacial free energy if

$$O^{ST}\sigma^{ST} + O^{TU}\sigma^{TU} + O^{TV}\sigma^{TV} < O^{TU}\sigma^{SU} + O^{TV}\sigma^{SV} \quad (84.5)$$

where  $O^{ST}$ , etc. are the surface areas and  $\sigma^{ST}$ , etc. are the specific free energies of the various interfaces.

When the inequality (84.5) holds, a small twin will persist and the boundaries moving away from the point of formation will lead to a macroscopic twinned region  $T$ . As the twin boundary energy will be very low when the twin interface is fully coherent, there is a reasonable probability of the inequality being valid. Suppose the boundaries  $S-U$  and  $T-V$  for example are both random high angle boundaries with nearly equal energies. Then the condition is

$$\sigma^{SU} - \sigma^{TV} > (O^{ST}/O^{TU})\sigma^{ST} \quad (84.6)$$

A sufficient difference in the left-hand side of this condition might be produced in two ways. The most obvious is when the  $S-U$  interface is a high angle boundary and the  $T-U$  interface is a low angle boundary, or an approximation to a twin boundary. The other possibility is that both boundaries are high angle but the  $T-U$  boundary is a special or coincident site boundary of low energy.

The Fullman-Fisher theory also applies to the situation shown in Fig. 19.12 where a new grain contact is established between grains  $S$  and  $V$ . If a twin has an orientation close to that of grain  $V$  then the  $T-V$  interface will have low energy and the configuration with the twin will have a lower energy than the configuration without it.

The Fullman-Fisher theory is supported by direct observations of twins at moving grain corners (Burke, 1949; Bolling and Winegard, 1958b). A parallel-sided twin which crosses a grain may be explained by a second growth accident which restores the original orientation  $S$ . This would be energetically unfavourable of course if  $T$  were still in contact with  $U$  and  $V$ , but could well occur at a later stage of recrystallization when  $T$  is growing into quite different grains. A possible sequence leading to the formation of a terminated

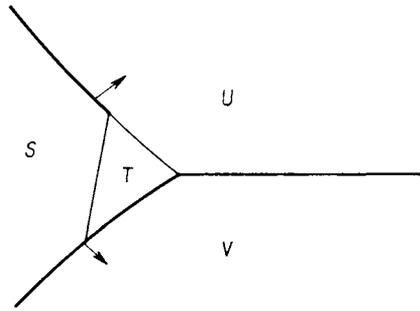


FIG. 19.11. The formation of an annealing twin at a three-grain junction (after Fullman and Fisher, 1951).

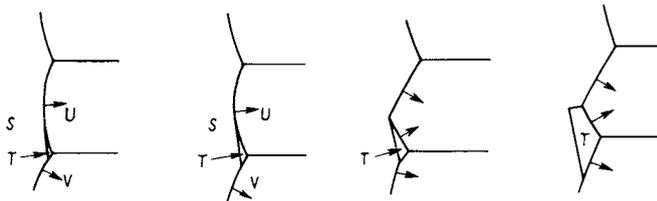


FIG. 19.12. Formation of a twin with a coherent and a non-coherent interface during grain boundary migration (after Fullman and Fisher, 1951).

twin is shown in Fig. 19.12. The coherent twin interface becomes tangential to the main migrating grain boundary, after which another interface must form.

Direct evidence for the formation of annealing twins during grain boundary motion, with the formation of a coincident site boundary, has been obtained in high purity lead by Aust and Rutter (1960b) using the technique described earlier. In some cases, the first  $T-U$  boundary had a rather low density of coincident sites, and a second twinning act produced a new orientation  $T'$  with a higher density of coincident sites in  $T'-U$ , and hence presumably a lower energy. Similar results were obtained by Aust (1961) for high purity aluminium. Because of the rather high energy of a twin boundary in this metal, the inequality is much harder to satisfy and only high density coincident site boundaries are effective in introducing twins. Annealing twins are comparatively rare in aluminium but are prolific in copper- and silver-based alloys of very low stacking fault (and hence twin boundary) energy.

No direct evidence has been reported for twin formation when the boundary  $T-U$  becomes a low angle boundary or another twin boundary. The probable reason for this is that although the situation is energetically favourable, the  $T-U$  boundary would then be relatively immobile, and growth of the twin would be impeded. Special boundaries are effective because they combine a low energy with a high mobility. However, Hu and Smith (1956) have shown that the number of twin lamellae per unit grain boundary area

depends only on the number of new grain contacts which have been made during growth, thus suggesting that the sequence shown in Fig. 19.12 is the most usual method of introducing a twin.

## REFERENCES

- ANDERSON, W. A. and MEHL, R. F. (1945) *Trans. Metall. Soc. AIME* **135**, 416.
- AUST, K. T. (1961) *Trans. Metall. Soc. AIME* **221**, 758; (1974) *Can. Metall. Q.* **13**, 133.
- AUST, K. T. and RUTTER, J. W. (1959a) *Trans. Metall. Soc. AIME* **215**, 119; (1959b) *Ibid.* **215**, 820; (1960a) *Ibid.* **218**, 50; (1960b) *Ibid.* **218**, 1023.
- BAILEY, J. E. (1960) *Philos. Mag.* **5**, 833.
- BAILEY, J. E. and HIRSCH, P. B. (1960) *Philos. Mag.* **5**, 485; (1962) *Proc. R. Soc. London, Ser. A* **267**, 11.
- BAINBRIDGE, D. W., LI, C. H., and EDWARDS, E. H. (1954) *Acta Metall.* **2**, 322.
- BECK, P. A. (1953) *Acta Metall.* **1**, 422; (1954) *Adv. Phys.* **3**, 245.
- BECK, P. A., KREMER, J. C., DEMER, L. J., and HOLZWORTH, M. L. (1948) *Trans. Metall. Soc. AIME* **175**, 372.
- BECK, P. A. and SPERRY, P. R. (1950) *J. Appl. Phys.* **21**, 150.
- BECK, P. A., SPERRY, P. R., and HU, H. (1950) *J. Appl. Phys.* **22**, 420.
- BEVER, M. B., HOLT, D. L., and TITCHENER, A. C. (1973) *Prog. Mater. Sci.* **17**, 1.
- BOLLING, G. F. and WINEGARD, W. C. (1958a) *Acta Metall.* **6**, 283; (1958b) *Ibid.* **6**, 288.
- BOLLMAN, W. (1958-9) *J. Inst. Met.* **87**, 439.
- BURGERS, W. G. (1947) *Proc. Acad. Sci. Amst.* **50**, 452.
- BURGERS, W. G. and TIEDMA, T. J. (1953) *Acta Metall.* **1**, 234.
- BURKE, J. E. (1949) *Trans. Metall. Soc. AIME* **180**, 73.
- BURKE, J. E. and TURNBULL, D. (1952) *Prog. Met. Phys.* **3**, 220.
- CAHN, J. W. (1962) *Acta Metall.* **10**, 789.
- CAHN, R. W. (1949) *J. Inst. Met.* **76**, 237; (1950) *Proc. Phys. Soc. London, Sect. A* **63**, 323.
- CLAREBOROUGH, L. M., HARGREAVES, M. E., and WEST, G. W. (1955) *Proc. R. Soc. London, Ser. A* **232**, 252; (1956) *Philos. Mag.* **1**, 528.
- COTTELL, A. H. (1953) *Dislocations and Plastic Flow in Crystals*, p. 188, Clarendon Press, Oxford.
- CRUSSARD, C. (1944) *Rev. Metall.* **41**, 118.
- DAVIES, P. W., GREENOUGH, A. P., and WILSHIRE, B. (1961) *Philos. Mag.* **6**, 795.
- DECKER, B. F. and HARKER, D. (1950) *Trans. Metall. Soc. AIME* **188**, 887.
- DOHERTY, R. D. (1982) *Met. Sci.* **16**, 1.
- FELTHAM, P. (1957) *Acta Metall.* **5**, 97.
- FROIS, C. and DIMITROV, O. (1961) *C.R. Acad. Sci.* **252**, 1465.
- FULLMAN, R. L. and FISHER, J. C. (1951) *J. Appl. Phys.* **22**, 1350.
- GAY, P., HIRSCH, P. B., and KELLY, A. (1954) *Acta Crystallogr.* **7**, 41.
- GAY, P. and KELLY, A. (1953) *Acta Crystallogr.* **6**, 165.
- GRAHAM, C. D. and CAHN, R. W. (1956) *Trans. Metall. Soc. AIME* **206**, 57.
- GREEN, R. E., LIEBMAN, B. G., and YOSHIDA, H. (1959) *Trans. Metall. Soc. AIME* **215**, 610.
- GUINIER, A. and TENNEVIN, J. (1948) *Acta Crystallogr.* **1**, 188.
- HIBBARD, W. A. and DUNNE, C. G. (1957) *Creep and Recovery*, American Society for Metals, Metals Park, OH.
- HIRSCH, P. B., HORNE, R. W., and WHELAN, M. J. (1957) *Dislocations and Mechanical Properties of Crystals*, p. 92, Wiley, New York.
- HOLMES, E. L. and WINEGARD, W. C. (1959a) *Acta Metall.* **7**, 411; (1959b) *Can. J. Phys.* **37**, 496; (1959-60) *J. Inst. Met.* **88**, 468.
- HU, H. (1959) *Trans. Metall. Soc. AIME* **215**, 320; (1962) *Acta Metall.* **10**, 1112; (1963) *Recovery and Recrystallization of Metals*, ed. L. Himmel, p. 311, Interscience, New York.
- HU, H. and SZIRMAI, A. (1961) *Trans. Metall. Soc. AIME* **221**, 839.
- JACQUET, P. A. (1954) *Acta Metall.* **2**, 752, 770.

- KELLAR, J. N., HIRSCH, P. B., and THORP, J. S. (1950) *Nature (London)* **156**, 554.
- KRONBERG, M. L. and WILSON, F. H. (1949) *Trans. Metall. Soc. AIME* **185**, 50.
- LACOMBE, P. and BERGHEZAN, A. (1949) *Physica* **15**, 161.
- LAMBOT, H. J., VASSAMILLET, L., and DEJACE, J. (1953) *Acta Metall.* **1**, 711.
- LESLIE, W. C., MICHLAK, J. T., and AUL, F. W. (1963) *Iron and its Dilute Solid Solutions*, p. 119, Interscience, New York.
- LI, C. H., EDWARDS, E. H., WASHBURN, J., and PARKER, E. R. (1953) *Acta Metall.* **1**, 223.
- LIEBMANN, B. and LÜCKE, K. (1956) *Trans. Metall. Soc. AIME* **206**, 1413.
- LIEBMANN, B., LÜCKE, K., and MASING, G. (1956) *Z. Metallkd.* **47**, 57.
- LÜCKE, K. and DETERT, K. (1957) *Acta Metall.* **5**, 628.
- MOTT, N. F. (1948) *Proc. Phys. Soc., London* **60**, 391.
- ORIANI, R. A. (1959) *Acta Metall.* **7**, 62.
- PARTHASARATHI, M. N. and BECK, P. A. (1961) *Trans. Metall. Soc. AIME* **221**, 831.
- RATH, B. B. and HU, H. (1972) *The Nature and Behaviour of Grain Boundaries*, ed. H. Hu, p. 405, Plenum Press, New York.
- RATHENAU, G. W. (1952) *Rep. 9th Solvay Conf. Solid State*, p. 55.
- RUTTER, J. W. and AUST, K. (1960) *Trans. Metall. Soc. AIME* **218**, 682.
- SEYMOUR, W. E. and HARKER, D. (1950) *Trans. Metall. Soc. AIME* **188**, 1001.
- SMITH, C. S. (1948) *Trans. Metall. Soc. AIME* **175**, 151.
- STANLEY, J. K. and MEHL, R. F. (1942) *Trans. Metall. Soc. AIME* **161**, 140.
- THORNTON, P. H. and CAHN, R. W. (1961) *J. Inst. Met.* **89**, 455.
- TIEDMA, T. J., MAY, W., and BURGERS, W. G. (1949) *Acta Crystallogr.* **2**, 151.
- TITCHENER, A. L. and BEVER, M. B. (1958) *Prog. Met. Phys.* **7**, 247.
- TURNBULL, D. (1951) *Trans. Metall. Soc. AIME* **191**, 661.
- VANDERMEER, R. A. and GORDON, P. (1959) *Trans. Metall. Soc. AIME* **215**, 517; (1963) *Recovery and Recrystallization of Metals*, p. 211, Interscience, New York.
- WALTER, J. L. (1959) *Acta Metall.* **7**, 424.
- WALTER, J. L. and DUNN, C. G. (1959) *Trans. Metall. Soc. AIME* **215**, 465.
- WARRINGTON, D. H. (1961) *Eur. Regional Conf. Electron Microscopy, Delft, Netherlands*, p. 354.
- WASHBURN, J. and PARKER, E. R. (1952) *Trans. Metall. Soc. AIME* **194**, 1076.
- WHEELAN, M. J., HIRSCH, P. B., HORNE, R. W., and BOLLMAN, W. (1957) *Proc. R. Soc. London, Ser. A* **240**, 524.
- WOOD, W. A. (1951) *Philos. Mag.* **42**, 310; (1953) *Bull. Inst. Met.* **1**, 198.

## CHAPTER 20

# *Deformation Twinning*

### 85. CRYSTALLOGRAPHY OF TWINNING

An outline description of the crystallography of deformation twinning was given in Section 8, where the twinning elements  $K_1$ ,  $K_2$ ,  $\eta_1$  and  $\eta_2$ , together with the plane of shear,  $S$ , and the shear magnitude,  $s$ , were defined. These elements define a *twinning mode* and this section will be concerned mainly with the factors which determine the modes observed in particular crystal structures. In the formal theory, no restriction will be placed on the twinning elements, and the criteria used to select the operative modes, such as reasonably small values of  $s$ , are partly intuitive and have to be justified by comparison with experiment. This approach is best for most twins, but there are some special examples in which the twin boundary may be regarded as a particular case of a high angle grain boundary.  $K_2$  and  $\eta_2$  must then be operative slip elements of the original (or parent)<sup>†</sup> structure.

The theory to be developed will include a more rigorous treatment of the two types of deformation twin discussed on p. 54 and of the orientation relations between parent and twin crystals. In the classical theory, two twinned structures are related by reflection in a plane (the twinning plane) or by a rotation of  $180^\circ$  about a direction (the twinning direction) and this leads to four different relative orientations in the most general case. As discussed on p. 54, a sufficient condition for a simple shear to reproduce the original lattice, or some superlattice of the original lattice, is that a unit cell of the structure, defined by three non-coplanar vectors in  $K_1$  and  $K_2$ , is sheared into an equivalent cell. The orientation of the new (twin) lattice relative to the original lattice may then be described either as a reflection in  $K_1$  or as a rotation of  $180^\circ$  about  $\eta_1$ . However, the important feature of deformation twinning is the shape change resulting from the simple shear, and it thus seems logical to consider the possibility of a more general definition which rests only

<sup>†</sup> Twinning is a mutual condition and each crystal of a twin pair may be regarded as the twin of the other. In deformation twinning, this reciprocity means that equal but oppositely directed shears will convert region 1 into region 2 and vice versa. In some experimental conditions, such as twin arrays produced by martensitic transformation, the twin boundaries may readily be induced to move in opposite directions by opposite external stresses. Often, however, deformation twins nucleate and grow from an original single crystal matrix which is then conventionally described as the parent structure. The nomenclature is a little unfortunate, inasmuch as the parent of a given twin is also a twin of that twin.

on the requirement that a simple shear deformation of the parent lattice will reproduce this lattice in another orientation. This more general definition, proposed by Crocker (1959) and Bilby and Crocker (1965), leads to the prediction of additional possible modes, in which either three or four of the twinning elements may be irrational, and to non-classical orientation relations between twin and parent.

As already noted on p. 54, there are only two classical orientation relations for lattices and for centrosymmetric structures, and these are often described simply as "reflection" and "rotation" twins respectively, it being understood that the type I reflection is in  $K_1$  and the type II rotation is about  $\eta_1$ . The division into type I and type II twins is most conveniently made on the basis of rational  $K_1$  and  $\eta_2$  or rational  $K_2$  and  $\eta_1$  respectively, but many structures form compound twins in which all four elements are rational. For compound twins, the plane of shear must be rational and, if it is a mirror plane of the structure, orientations I and II become identical (see below). However, there is no general requirement of this type so that, even with compound twins, it may be necessary to distinguish between two possible orientation relations which may be called type I compound and type II compound twins respectively (Rowlands *et al.*, 1968). For clarity, Christian and Laughlin (1988) have suggested that when the type I and type II orientations of a compound mode are equivalent, it should be described as having a "combined" orientation.

The simple crystallography of compound twins in structures of high symmetry led to a method of deducing the twinning elements by choosing a plane of symmetry normal to  $K_1$  as the plane of shear, thus defining  $\eta_1$ . A procedure of this kind has no obvious physical significance and also has the disadvantage of resting on an experimental determination of  $K_1$ , whereas the aim of a theory should be to predict all of the twinning elements. The method does give the correct twinning elements for many of the twins observed in metals (Hall, 1954), but becomes unwieldy or incorrect in cases where some of the elements are irrational. A more systematic theory was developed by Jaswon and Dove (1956, 1957, 1960), who assumed that the twinning elements may be selected by minimizing the magnitude of the shear. They considered first a type I twin in a simple Bravais lattice, so that parent and twin are mirror images in  $K_1$ . If the twinning shear is to move all the parent lattice sites to their correct twin positions, the possible  $\eta_1$  directions are readily found by projecting parent sites onto the common  $K_1$  plane.

Consider a set of parallel lattice planes of spacing  $d$  and label them consecutively as ...  $\bar{3}\bar{2}\bar{1}0123$ ..., etc. A homogeneous shear  $s$  of the lattice on the positive side of plane 0 is now specified by translating planes 123... parallel to themselves through distances  $sd$ ,  $2sd$ ,  $3sd$ , etc. in the direction of the unit vector  $l$  parallel to  $\eta_1$ . For a type I twin in which *all* lattice points are moved to their final positions by the shear, the new positions of the lattice points must be mirror images in  $K_1$  of their original positions. Thus if the lattice sites of plane  $\bar{1}$  are projected onto plane 1, the possible  $\eta_1$  directions are given by the set of vectors connecting any site of plane 1 to any site projected from plane  $\bar{1}$ .

We have seen previously in Fig. 8.12 that the assumption that the shear moves all the lattice points to their final positions is not necessarily correct, as the structure of the interface may require in addition a relative translation of the parent and sheared lattices.

However, any such translation  $\mathbf{t}$  [see eqn. (9.1)] cannot be detected macroscopically or by X-ray diffraction, so that the classical descriptions of the orientation relations remain valid. For the purpose of deducing the crystallographic elements of a particular twinning mode, only the shear components of the total relative displacements need be considered and neither non-zero values of  $\mathbf{t}$  nor any inhomogeneous displacements in the interface region will affect the calculation of the mode of minimum shear. At the atomic level, however, such a translation destroys the pseudosymmetry (or “antisymmetry”) operations (Pond, 1985, 1988; Pond and Vlachavas, 1983) which relate parent and twin structures; in Fig. 8.12, for example, the reflection symmetry of configuration (a) is destroyed in (b). There are further complications in the descriptions of the orientation relations in non-symmorphic structures which exhibit mirror glide planes or screw rotation axes connecting identical atoms in non-equivalent lattice positions. For example, if the plane of shear is a mirror glide plane in such a structure, the classical type I and type II twinning orientations are in principle distinct but differ from each other only by a rigid translation, so that they are equivalent if the only measurements are of the relative orientations of twin and parent. A further discussion of  $\mathbf{t}$  will be given in relation to interface structures and defects. Of course it is quite likely that the operative mode is determined by the interface energy and hence by its atomic structure rather than simply by the value of  $s$ , but that was not Jaswon and Dove’s hypothesis.

Any plane of the parent lattice which is not itself a mirror plane could serve as the  $K_1$  plane, but it is easy to select the plane which minimizes the lattice shear. Figure 20.1 shows two lattice sites,  $P$  and  $R$ , in planes  $l$  and  $\bar{l}$ , together with the projection  $R'$  of  $R$  in  $l$  and it follows that

$$s^2 = (u^2/d^2) - 4 \quad (85.1)$$

where  $\mathbf{u}$  is the vector displacement  $P \rightarrow R$ . Thus a small shear requires a large interplanar spacing and a small vector  $\mathbf{u}$  connecting sites in planes  $l$  and  $\bar{l}$ . Equation (85.1) may be used directly to find the minimum shear by inspection, but Jaswon and Dove found it convenient to write  $u \geq b$ , where  $\mathbf{b}$  is the smallest possible lattice vector. The resulting inequality

$$d^2 \geq b^2/(s^2 + 4) \quad (85.2)$$

may be regarded as a condition restricting the possible values of  $d$  for any chosen maximum value of  $s$ . Thus by setting  $s = 1$ , for example, the planes for which the shear is

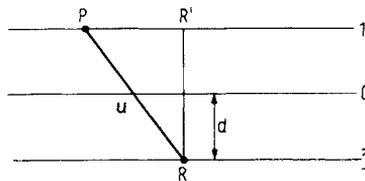


FIG. 20.1. Determination of the shear in type I twinning (after Jaswon and Dove, 1956).

less than unity may readily be enumerated and tested. Jaswon and Dove assumed that the actual  $K_1$  plane will be that plane, which is not a mirror plane, for which the shear is least. Clearly, if there are no planes satisfying the inequality within the chosen maximum value of  $s$ , a higher value must be selected. An equivalent procedure is to calculate  $s$  for the closest-packed planes (largest  $d$ ) which are not mirror planes. Substitution of this value of  $s$  in the inequality (85.2) will then show which other planes, if any, need to be considered.

Jaswon and Dove's procedure easily leads to the prediction of the actual twinning modes of almost all of the metallic structures in which the atoms occupy the sites of a single Bravais lattice. For example, in the b.c.c. structure,  $b^2 = 3a^2/4$  and, for a shear of less than unity,  $d^2$  must be greater than  $3a^2/20$ . The only planes satisfying the inequality are  $\{110\}$ ,  $\{100\}$  and  $\{112\}$ . The first two planes are mirror planes so that the  $\{112\}$  planes are the only possible  $K_1$  planes if the shear is to be less than unity. The shear then has magnitude  $s = 2^{-1/2}$  and the  $\eta_1$  and  $\eta_2$  directions are  $\langle \bar{1}\bar{1}1 \rangle$  and  $\langle 111 \rangle$  respectively. Similarly, for the f.c.c. structure,  $b^2 = a^2/2$  and, if  $s$  is less than unity,  $d^2$  must be greater than  $a^2/10$ . The planes satisfying the inequality are now  $\{111\}$ ,  $\{100\}$  and  $\{110\}$  and thus only the  $\{111\}$  planes are possible  $K_1$  planes within this restricted range of  $s$ . In contrast to the b.c.c. structure, the  $K_1$  planes are thus the closest-packed planes of the f.c.c. structure. In this case  $s$  is again  $2^{-1/2}$  whilst  $\eta_1$  and  $\eta_2$  are  $\langle 11\bar{2} \rangle$  and  $\langle 112 \rangle$  respectively.

For tetragonal structures, the inequality theorem indicates that the  $K_1$  plane is  $\{101\}$  and the  $\eta_1$  direction is  $\langle 10\bar{1} \rangle$  for all axial ratios  $\gamma = c/a$  which are near to unity. This corresponds to the observed twinning mode in indium which has a face-centred tetragonal (f.c.t.) structure with  $\gamma \cong 1.08$ . The mode has  $s = \gamma - \gamma^{-1}$  and has an interesting relationship with the two cubic twinning modes. The f.c.t. structure may equivalently be regarded as body-centred tetragonal (b.c.t.) and it reduces to f.c.c. when  $\gamma_{\text{fct}} = 1$  ( $\gamma_{\text{bct}} = 2^{-1/2}$ ) and to b.c.c. when  $\gamma_{\text{fct}} = 2^{-1/2}$  ( $\gamma_{\text{bct}} = 1$ ). The f.c.t.  $\{101\}$  twinning mode has  $s = 0$  (and thus vanishes) at the first limit, and it becomes equivalent to the b.c.c.  $\{112\}$  mode with  $s = 2^{-1/2}$  at the second limit. In contrast, the b.c.t.  $\{101\}$  mode has  $s = 0$  at the second limit and becomes equivalent to the f.c.c.  $\{111\}$  mode with  $s = 2^{-1/2}$  at the first limit.

Some alloys, and especially many steel martensites, are known to have tetragonal structures for which  $\gamma_{\text{bct}}$  is close to unity. Transformation twinning of  $\{101\}$  type has been observed in b.c.t. gold manganese alloys (Smith and Gaunt, 1962) but both transformation and deformation twins in steel martensites usually have  $K_1 = \{112\}$  of the b.c.t. lattice, so that in this case the dominant twinning mode is derived from the b.c.c. mode and the predictions of the minimum shear hypothesis are not fulfilled. Another example of the failure of this hypothesis in single lattice structures is provided by crystalline mercury. If the structure is referred to a face-centred rhombohedral lattice, the mode of lowest shear is compound with conjugate  $K_1$  planes of type  $\{110\}$  and  $\{001\}$ , but early work in which  $\{110\}$  twins were identified now seems to have been incorrect and the operative mode is actually a type II twin with a larger shear.

The Jaswon-Dove approach can also be used for type I twins in structures containing two atoms per primitive unit cell (so-called double lattice structures). The simplest assumption in this case is that the shear moves each lattice point into its mirror image

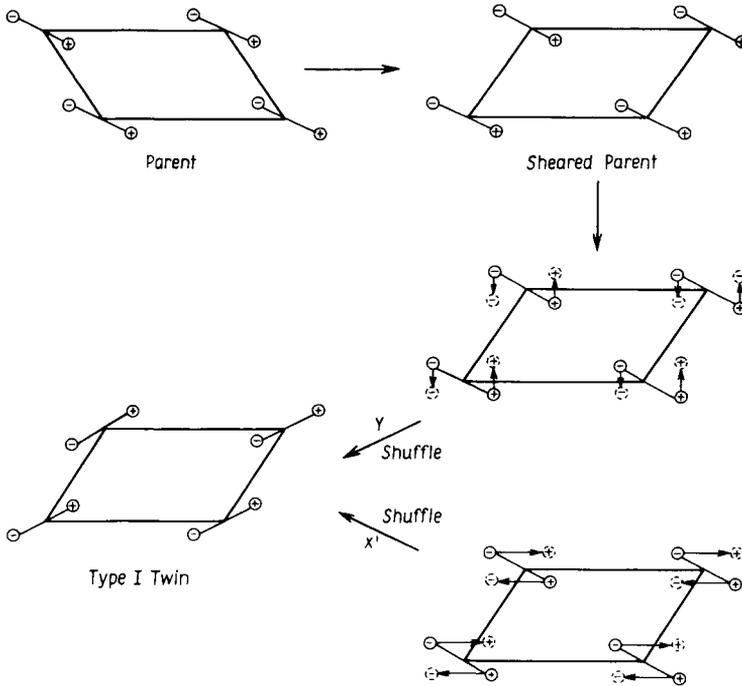


FIG. 20.2. Diagram illustrating possible twinning mechanisms in a double lattice structure. Circles marked + and - represent atoms at equal distances above and below the plane of projection. The sheared parent shows the structure after applying a homogeneous simple shear to all the lattice points, and the shuffles marked  $X'$  and  $Y$  represent the  $X$  and  $Y$  mechanisms respectively of Jaswon and Dove (1956). In the  $Y$  mechanism, the atom displacements are normal to  $K_1$ , whereas in the  $X'$  mechanism the atom movements must be in this plane but not necessarily parallel to  $\eta_1$ .

position in the interface, but that the motif unit of two atoms associated with the lattice point is treated as a rigid unit during the shear. Additional displacements of the atoms are then generally necessary, and these are described as "shuffles" because, unlike the shear, they produce no macroscopic effects. For a double lattice structure, there are two simple shuffles leading to a type I twin, as illustrated in Fig. 20.2, and, following Jaswon and Dove, these are described as the  $X'$  and  $Y$  mechanisms respectively. ( $X'$  is used rather than  $X$  in order to distinguish it from another mechanism, shown in Fig. 20.3, which Jaswon and Dove also called  $X$ .) Note that if the  $K_1$  lattice planes are regarded as corrugated atomic planes, the shuffles reverse the asymmetric corrugations.

This assumption about the division of the net atomic displacements into shear and shuffle components is clearly somewhat arbitrary and it might alternatively be assumed that the shuffles precede the shear. Moreover, in the description given by Jaswon and Dove and later followed by Bilby and Crocker, the motif units are treated as rigid during the shear, so that each atom of a double lattice structure is given the shear displacement appropriate to the midpoint of the motif pair. An equally valid factorization of the net

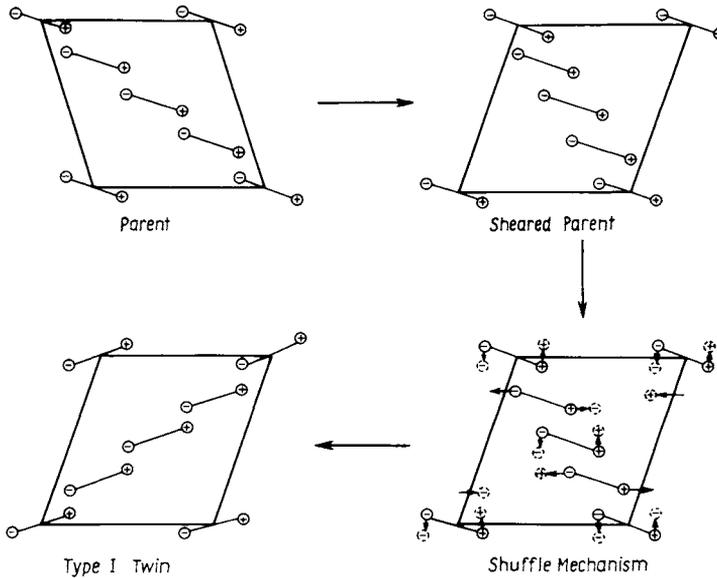


FIG. 20.3. Diagram illustrating the  $X$  shuffling mechanism of Jaswon and Dove (1957). Circles marked + and - represent atoms at equal distances above and below the plane of projection. The sheared parent shows the structure after a shear which moves one-half of the lattice points to their new positions; this arises because a primitive lattice vector in the  $\eta_2$  direction crosses four lattice planes ( $q=4$ ). The atom displacements in successive  $K_1$  planes are alternately normal and parallel to  $K_1$ .

movements would be first to give each atom the displacement of the homogeneous shear (i.e. to treat the atoms as embedded in a homogeneously sheared continuum) and then to define the shuffles as the remaining atomic displacements. Whether or not the alternative descriptions have any physical significance depends on the models used for interface structure and migration. There may also be a difference in the formal criteria for multiple lattice structures to be able to form twins without shuffles (see p. 865).

The previous notation is now extended so that successive *atomic* planes are denoted  $\dots 2\bar{a} \bar{2}b \bar{1}\bar{a} \bar{1}b 0a 0b 1a 1b 2a 2b \dots$ , and the Jaswon and Dove treatment requires that the shear gives planes  $1a$  and  $1b$  an identical parallel displacement which brings the atoms in these planes immediately over the atoms in  $1\bar{a}$  and  $1\bar{b}$  respectively. In the  $Y$  mechanism, the atoms are also given further shuffle displacements normal to  $K_1$  which effectively interchange the  $1a$  and  $1b$  planes so that they become mirror images of the  $1\bar{b}$  and  $1\bar{a}$  planes, thus producing the twin. (If the motif unit is not regarded as rigid during the shear, the shuffle displacements in the  $Y$  mechanism will be parallel to the final  $\eta_2$  direction [i.e.  $\eta'_2$  of Fig. 2.4] if the shuffles are supposed to follow the shear and to the original  $\eta_2$  direction if the shuffles precede the shear.) In the  $X'$  mechanism, the atoms are given shuffle displacements within the  $K_1$  planes to give the same result. Note that the shear alone cannot describe the twinning because the crystal structure produced by the

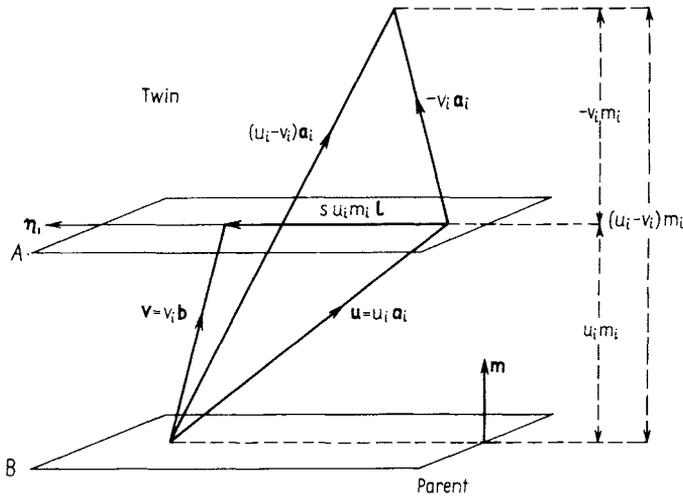


FIG. 20.4. Crystallographic relations in type I twinning (after Crocker, 1959). When the interface moves from position  $A$  to  $B$ , the vector  $u$  in the parent becomes  $v$  in the twin. The components of  $u$  are  $u_i$  in the parent basis  $a_j$  and those of  $v$  are  $v_j$  in the twin basis  $b_j$ . The Figure shows that the projections of the vectors  $u_i a_i$  and  $v_i b_i$  on to the normal to the  $K_1$  plane are both equal to one-half of the projection of  $(u_i - v_i) a_i$  on to  $m$ .

shear (Fig. 20.2) is different from that of the original (parent) phase. No reference has been made to the structure of the interface region (i.e. to the atoms originally in planes  $0a$  and  $0b$ ) in describing either the shear or the shuffles. One obvious possibility is that these atoms are given half the displacements of the  $Y$  mechanism to produce a flat atomic interface midway between the opposite corrugations in the lattice planes of parent and twin, and this is believed to be a good model for some h.c.p. twin interfaces. However, the macroscopic effects of the twinning are independent of the atomic configuration at the interface, although this structure is of considerable importance in growth mechanisms and is discussed further in Section 87.

The two types of shuffle shown in Fig. 20.2 are the only possibilities for double lattice structures in which all lattice sites are translated to their correct positions by the macroscopic shear. If the double lattice structure is regarded as a set of motif units situated at the points of a lattice, neither shuffle disrupts the motif pairs. However, the necessity for atomic shuffles lessens the significance of shearing all lattice sites to their correct positions and, in metallic structures at least, the selected motif units are usually without appreciable physical significance (i.e. they do not represent molecules). Thus it seems reasonable to examine the possibility of twinning modes in which the macroscopic shear carries only a fraction of the sites to their correct twinned positions, even if the subsequent atomic shuffles involve disruption of the motif units.

Jaswon and Dove extended their theory in this way to include cases in which only half of the parent lattice sites are moved to their final positions by the twinning shear. This led them in particular to consider the  $X$  shuffle mechanism shown in Fig. 20.3, in which the

atomic shuffles in successive planes are alternately parallel and normal to  $K_1$ , those parallel to  $K_1$  involving disruption of motif units.

The mechanisms of either Fig. 20.2 or Fig. 20.3 require that planes  $2a2b$  are carried over planes  $\bar{2}\bar{a}\bar{2}\bar{b}$  by the macroscopic shear. To investigate the possible twinning planes for this type of structure, it is thus only necessary to apply the previous projection technique to this pair of planes. An inequality analogous to (85.2) is obtained,

$$d^2 \geq b^2/4(s^2 + 4) \quad (85.3)$$

giving a necessary condition on the spacing of the  $K_1$  planes for any prescribed value of  $s$ . Clearly the condition on  $d^2$  can be still further relaxed by allowing smaller fractions of the parent sites to be sheared directly to the twin sites, at the expense of more complex shuffles, but consideration of this is deferred until later in this chapter.

In applying the theory to predict operative twinning modes in double lattice structures, it is now assumed that the mode selected will be that giving the smallest shear, irrespective of whether this involves an  $X$  or  $Y$  type shuffle. If the shears are comparable for alternative modes involving  $X$  or  $Y$  shuffles, however, the relative probability of the two types of shuffle may be taken into consideration. For example, if the vector joining the two atoms of the motif unit is nearly parallel to  $K_1$ ,  $Y$  shuffles will require only very small atomic displacements and hence may be considered more probable than  $X$  shuffles.

Although the Jaswon and Dove theory has been described only for type I reflection twins, essentially similar procedures may also be used for type II rotation twins (Jaswon and Dove, 1960). Instead of considering the displacements parallel to  $K_1$ , those parallel to the rational  $K_2$  plane are now examined. The theory as a whole is very successful in predicting the operative twinning modes in single lattice structures and in most double lattice structures, including the most common h.c.p. mode, the modes observed in bismuth, arsenic and antimony, the mode usually observed in diamond structures, and both type I and type II twins in  $\alpha$ -uranium. However, the theory does not explain why several modes are observed in some structures, nor does it predict correctly their relative frequency, and some observations (e.g. "anomalous" h.c.p. modes) are unexplained. A description based entirely on the twinning shear cannot distinguish between a mode and its conjugate, but when the  $K_1$  and  $K_2$  planes are not crystallographically equivalent, it is frequently found that only one of these modes is observed. Finally, the theory does not consider non-centrosymmetric structures where there may be four different orientation relations.

The suggestion that the magnitude of the twinning shear is an important factor in determining the operative twinning mode or modes was also made by Kiho (1954, 1958), whose first paper predated that of Jaswon and Dove. He considered specifically the atom movements at an idealized parent-twin interface, and assumed that each atom moves to the nearest available twin site, and that the vector sum of the shuffles is zero. The shuffles which he described included the  $X$  and  $Y$  mechanisms, together with another mechanism to explain the anomalous twins in titanium, but he did not give a full treatment of the uranium modes. Kiho also suggested that in choosing between a twinning mode and its conjugate, the mode for which the Burgers vector of a twinning dislocation in the interface is least

should be preferred. This is equivalent to a statement that the preferred mode of a conjugate pair should be that for which the spacing of the lattice  $K_1$  planes is smaller.

The available experimental results undoubtedly show that, in some cases, the magnitude of the shear is not the only factor which controls the operative twinning modes. The theory of Bilby and Crocker (1965), which includes a more rigorous treatment of the orientation relations and shuffle mechanisms, will now be described.

As a parent crystal and its twin remain in contact at the interface plane during the formation of the twin, it is clear that the relation between the structures must be such that this plane is invariant in any deformation relating the two lattices. This is automatically accomplished in the shear description but, in specifying the orientation relations, consideration is given to proper or improper rotations which will carry one lattice into the other. Consideration of the operations of this type which will leave the  $K_1$  plane unaltered leads at once to the four orientation relations mentioned on p. 51, namely:

- (I) reflection in  $K_1$ ;
- (II) rotation of  $180^\circ$  about  $\eta_1$ ;
- (III) reflection in the plane normal to  $\eta_1$ ; and
- (IV) rotation of  $180^\circ$  about the direction normal to  $K_1$ .

It is also possible to have two orientations of a structure which do not correspond to any of the relations I–IV, but in which the lattices (or suitable superlattices) are connected by a simple shear. If such a shear is regarded as a form of deformation twinning, as in the Bilby Crocker formulation mentioned on p. 860, the classical definition of twinning in terms of symmetry operations is inadequate. In fact, new non-classical twinning modes were first considered to arise from the combination of two twinning operations of the classical kind, a process which has been termed “double twinning” (Crocker, 1962). A more systematic theory of twinning as a shear process was subsequently developed by Bevis and Crocker (1968, 1969), and this enables the possible classical and non-classical modes with shear magnitudes smaller than any fixed value to be enumerated. However, there is currently no very convincing experimental evidence for the occurrence of non-classical twinning, so that the assumption will first be made that the classical orientation relations are valid and a discussion of the general theory is deferred to the end of this section. The orientation relations I–IV, and the associated division into type I and type II twins, follow necessarily from the more general shear definition if the assumption is made (as on p. 54) that there exists a cell of the parent which shears into an equivalent cell of the twin.

Let the parent lattice be defined by the three non-coplanar vectors  $\mathbf{a}_i$  forming the basis A, and the twin lattice by the vectors  $\mathbf{b}_i$  forming the basis B. These bases are chosen so that each set of vectors defines a similarly shaped *primitive* unit cell. One of the above four operations generates B from A, so that the four orientation relations may be written

$$\mathbf{b}_i^{(1)} = \mathbf{a}_i - 2(\mathbf{a}_i \cdot \mathbf{m})\mathbf{m},$$

$$\mathbf{b}_i^{(11)} = 2(\mathbf{a}_i \cdot \mathbf{l})\mathbf{l} - \mathbf{a}_i,$$

$$\begin{aligned} \mathbf{b}_i^{(III)} &= \mathbf{a}_i - 2(\mathbf{a}_i \cdot \mathbf{l})\mathbf{l}, \\ \mathbf{b}_i^{(IV)} &= 2(\mathbf{a}_i \cdot \mathbf{m})\mathbf{m} - \mathbf{a}_i, \end{aligned} \quad (85.4)$$

where  $\mathbf{m}$  and  $\mathbf{l}$  are unit vectors normal to the  $K_1$  plane and in the  $\boldsymbol{\eta}_1$  direction respectively. As  $\mathbf{b}_i^{(I)} = -\mathbf{b}_i^{(IV)}$  and  $\mathbf{b}_i^{(II)} = -\mathbf{b}_i^{(III)}$ , the lattices given by orientations I and IV are identical, as are those given by II and III. When the atomic positions are considered, however, the two orientations in each pair are seen to be equivalent only for structures which have a centre of symmetry. It is thus sufficient to consider only  $\mathbf{b}_i^{(I)}$  and  $\mathbf{b}_i^{(II)}$  when there are no more than two (identical) atoms per primitive unit cell (single or double lattice structures), but the other two relations may be needed for more complex structures. Although these do not occur in most metals known to twin, the theory will be developed in general form as far as is practicable.

Twin orientations I and II are readily seen to be related to each other by a reflection in the plane of shear. Let any parent vector have components  $x_i$  in an orthonormal basis defined by  $\mathbf{l}$ ,  $\mathbf{m}$  and the unit normal to the plane of shear,  $\mathbf{l} \wedge \mathbf{m}$ . Then for orientation I, its twin vector has components  $[x_1, -x_2, x_3]$  and for orientation II, it has components  $[x_1, -x_2, -x_3]$ . It follows that there is no distinction between orientations I and II if the two twin vectors are crystallographically equivalent, i.e. if the plane of shear is a mirror plane.

Consider a lattice point of the parent lying within the twin interface and identified with respect to an arbitrary origin by the parent vector  $\mathbf{u}$ . Let the interface move into the parent until it contains the origin, so that the lattice point is displaced through a distance  $s(\mathbf{u} \cdot \mathbf{m})$  in the  $\boldsymbol{\eta}_1$  direction. The vector  $\mathbf{u}$  thus becomes a new vector  $\mathbf{v}$ , where  $\mathbf{v} = \mathbf{u} + s(\mathbf{u} \cdot \mathbf{m})\mathbf{l}$ . We now let  $u_i$  be the (rational) components of  $\mathbf{u}$  in the basis A and  $v_i$  be the (rational) components of  $\mathbf{v}$  in the basis B.<sup>\*</sup> The vector  $u_i \mathbf{a}_i$  in the parent thus becomes  $v_i \mathbf{b}_i$  in the twin, where

$$u_i \mathbf{a}_i + su_i m_i \mathbf{l} = v_i \mathbf{b}_i. \quad (85.5)$$

Any of the relations (85.4) may now be substituted for  $\mathbf{b}_i$  in this equation. At present the concern is only with lattices, and use of  $\mathbf{b}_i^{(I)}$  and  $\mathbf{b}_i^{(II)}$  gives respectively

$$(u_i - v_i) \mathbf{a}_i = -su_i m_i \mathbf{l} - 2v_i m_i \mathbf{m} \quad (85.6a)$$

and

$$(u_i + v_i) \mathbf{a}_i = -su_i m_i \mathbf{l} + 2v_i l_i \mathbf{l} \quad (85.6b)$$

Taking the scalar product of both sides of these equations with  $\mathbf{m} = m_i \mathbf{a}_i^*$  gives

$$(u_i + v_i) m_i = 0, \quad (85.7)$$

<sup>\*</sup>The vector  $\mathbf{u}$  may have irrational components in the basis B, as may  $\mathbf{v}$  in the basis A. In this section, it will not be necessary to refer a single vector to the two different bases A and B, but use will be made of *different* vectors such as  $u_i^\Lambda \mathbf{a}_i$  and  $u_i^\mathbf{B} \mathbf{b}_i$  which have *the same* rational components in A and B respectively. For this reason, the identifying superscripts will be omitted from the symbols  $u_i^\Lambda$ ,  $v_i^\mathbf{B}$ , etc. In the following description, combined rational indices  $[u_i \pm v_i]$  correspond to indices  $[x' - y']$  of Bilby and Crocker because of a change of sign in eqn. (85.5).

in both cases. Hence for both orientations I and II, the rational lattice vector  $(u_i + v_i)\mathbf{a}_i$  must lie in the  $K_1$  plane. This development proves that the twinning plane must contain at least one rational lattice vector. Equation (85.6b) shows, moreover, that for orientation II this rational lattice vector is parallel to  $\mathbf{l}$ , i.e. to  $\boldsymbol{\eta}_1$ . For orientation I, eqn. (85.7) may be used to rewrite eqn. (85.6a) in the form

$$(u_i + v_i)\mathbf{a}_i = su_i m_i \mathbf{l} + 2(u_i \mathbf{a}_i - u_i m_i m_j \mathbf{a}_j). \quad (85.8)$$

The right-hand side of eqn. (85.8) consists of the sum of a vector parallel to  $\mathbf{l}$  and the projection of the vector  $2\mathbf{u}$  on the  $K_1$  plane. Thus the rational lattice vector  $(u_i + v_i)\mathbf{a}_i$  will vary as  $\mathbf{u}$  varies, and the  $K_1$  plane contains an infinite number of rational directions. This proves that, for orientation I,  $K_1$  is rational.

The above analysis can be repeated for orientations III and IV, in which case the rational vector in the interface is  $(u_i - v_i)\mathbf{a}_i$  and is parallel to  $\mathbf{l}$  for orientation III and represents an infinite set of vectors in  $K_1$  for orientation IV. Hence it has been shown that for orientations I and IV,  $K_1$  must be rational (type I twinning), and for orientations II and III,  $\boldsymbol{\eta}_1$  must be rational (type II twinning). Consider type I twinning first. Then from eqn. (85.7)

$$2u_i m_i = -2v_i m_i = (u_i - v_i)m_i. \quad (85.9)$$

This relation is illustrated in Fig. 20.4. The rational vector  $(u_i - v_i)\mathbf{a}_i$  of the parent lies in the plane containing  $\boldsymbol{\eta}_1$  and  $\mathbf{m}$ , and the projections of the vectors  $u_i \mathbf{a}_i$  and  $-v_i \mathbf{a}_i$  on  $\mathbf{m}$  are both equal to one-half of the projection of  $(u_i - v_i)\mathbf{a}_i$  on  $\mathbf{m}$ . Substituting into eqn. (85.6a) and using eqn. (85.4) now gives

$$(u_i - v_i)\mathbf{a}_i + s(u_i - v_i)m_i \mathbf{l} = -(u_i - v_i)\mathbf{b}_i \quad (85.10)$$

This equation shows that the twinning operation converts the lattice vector  $(u_i - v_i)\mathbf{a}_i$  into its own twin  $-(u_i - v_i)\mathbf{b}_i$ . It is easy to prove that no other vectors have this property, so that  $(u_i - v_i)\mathbf{a}_i$  defines a unique direction (the  $\boldsymbol{\eta}_2$  direction). From eqns. (85.6a) and (85.9), the twinning shear is given by the vector

$$s\mathbf{l} = -2 \frac{(u_i - v_i)\mathbf{a}_i}{(u_i - v_i)m_i} - \mathbf{m}. \quad (85.11)$$

The meaning of this equation is shown in Fig. 20.5. Let  $\mathbf{g}$  be a unit vector in the direction of  $\boldsymbol{\eta}_2$ , i.e. of  $(u_i - v_i)\mathbf{a}_i$ . Then eqn. (85.11) may be written

$$s\mathbf{l} = -2 \frac{\mathbf{g}}{\mathbf{g} \cdot \mathbf{m}} - \mathbf{m}. \quad (85.12)$$

and the magnitude of the shear is

$$s^2 = 4[(1/\mathbf{g} \cdot \mathbf{m})^2 - 1] \quad (85.13)$$

or, in terms of the angle  $2\phi$  shown in Figs. 2.4 and 20.5,

$$s = 2 \cot 2\phi$$

which is eqn. (8.2).

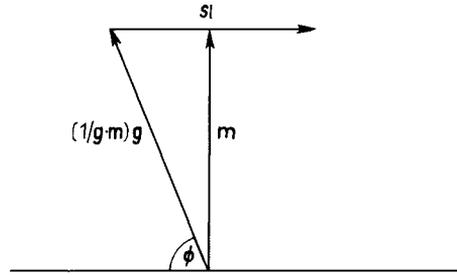


FIG. 20.5. The magnitude of the twinning shear. Unit vectors in the  $\eta_1$  direction, the direction normal to the  $K_1$  plane and the  $\eta_2$  direction are  $\mathbf{l}$ ,  $\mathbf{m}$  and  $\mathbf{g}$  respectively.

It has thus been proved that in type I twinning the  $K_1$  plane is rational, and there is a unique rational direction (the  $\eta_2$  direction) which is sheared into its own twin. The result has been deduced without making the assumption given on p. 54 that a unit cell of the parent is sheared into an equivalent cell of the twin. The elements  $K_1$  and  $\eta_2$  enable the amount of shear to be calculated, so that these two elements define the twin mode completely.

Next consider whether or not all the parent lattice points are carried to twin lattice sites by the shear. Let  $\mathbf{w} = w_i \mathbf{a}_i$  be a primitive lattice vector in the  $\eta_2$  direction, and let its projection along the normal to the  $K_1$  plane have magnitude

$$w_i m_i = qd \quad (85.14)$$

where  $q$  is a positive integer giving the number of lattice  $K_1$  planes traversed by this vector. The above discussion shows that all the lattice sites of the parent structure on the  $K_1$  plane defined by  $\mathbf{w}$  will become correctly positioned lattice sites of the twin structure as a result of the shear, but it is necessary (for  $q > 2$ ) to consider further the lattice sites on the  $(q - 1)$  intermediate planes. Note that whatever the structure of the interface, it reaches an equivalent position after moving forward a distance  $qd$ , so that the atom displacements are repeated in each successive group of  $q$  planes, and only one such group need be considered.

Let any lattice point in the nearest  $K_1$  plane to that through the origin be  $c_i \mathbf{a}_i$  (i.e.  $c_i m_i = d$ ). Bilby and Crocker assumed that, in type I twinning, all the lattice points on a given  $K_1$  plane shuffle in the same way, so that it is necessary to consider only one lattice point on each plane. Thus all the parent sites within the  $(q - 1)$  planes of interest are represented by  $pc_i \mathbf{a}_i$ , where  $p$  is a positive integer which is smaller than  $q$ . After the homogeneous shear, the positions of these sites will become  $pc_i \mathbf{a}_i + spd \mathbf{l}$  and, using eqns. (85.6a) and (85.9) with the above definitions of  $\mathbf{w}$  and  $p$ , this becomes

$$[pc_i - (2p/q)w_i] \mathbf{a}_i + 2pd \mathbf{m} \quad (85.15)$$

Now consider the twin sites. Any site in the  $q$ th  $K_1$  plane (defined by  $\mathbf{w}$ ) may be written as  $z_i \mathbf{b}_i$ , where from eqns. (85.10) and (85.14)

$$-z_i m_i = qd. \quad (85.16)$$

As the bases A and B define similar unit cells, all the twin lattice sites in the  $q - 1$  planes of interest are represented by the vectors

$$(z_i + rc_i)\mathbf{b}_i \quad (85.17)$$

where  $r$  (like  $p$ ) is a positive integer smaller than  $q$ . (Note that it is not necessary to consider  $p = q$  as we have shown that sites on this plane do not shuffle.)

The lattice shuffles must relate the twin lattice sites of eqn. (85.17) to the parent lattice sites of eqn. (85.15), and so are described by the vectors  $\Delta^1$ , obtained by subtracting eqn. (85.15) from eqn. (85.17). However, eqns. (85.4), (85.10) and (85.16) show that, for orientation I,

$$c_i\mathbf{b}_i = c_i\mathbf{a}_i - 2d\mathbf{m}$$

and

$$z_i\mathbf{b}_i = z_i\mathbf{a}_i + 2qd\mathbf{m}$$

so that after subtracting,

$$\Delta^1 = [(r - p)c_i + (2p/q)w_i + z_i]\mathbf{a}_i + 2d(q - r - p)\mathbf{m}. \quad (85.18)$$

For a given parent site, the parameters  $z_i$  and  $r$  representing the twin site may be regarded as disposable; that is, which parent and twin sites are related by the vector  $\Delta^1$  may be freely chosen. From eqn. (85.18) it follows that  $\Delta^1 \cdot \mathbf{m} = (q - p - r)d$ . Thus the component of  $\Delta^1$  normal to  $K_1$  is zero, i.e. the related parent and twin sites lie in the same  $K_1$  plane if  $p + r = q$ . In general, the shuffles might be expected to minimize  $|\Delta^1|$ . Although it is not always possible to choose the parameters so that  $\Delta^1 = 0$  for all sites, this can be done when  $q = 1$  or 2. For  $q = 1$ , there are no intermediate lattice sites to be shuffled and, for  $q = 2$ , there are only the sites on the plane defined by  $p = 1$ . The shuffle vectors for this plane are all zero if  $r = 1$  and  $z_i = -w_i$  are chosen. More generally, if  $q$  is even, the lattice points in the plane  $p = q/2$ , as well as those in the plane  $p = q$ , are sheared directly to their twin positions, as may be seen by choosing  $r = p$  and  $z_i = -w_i$ .

The case  $q = 2$  corresponds to Fig. 20.1, where it was assumed that the  $\eta_2$  direction relates lattice points on  $K_1$  planes a distance  $2d$  apart. All lattice points are then translated to their twin sites by the shear, as has just been shown analytically. When  $q = 1$ , there is a degenerate case of Fig. 20.1, in which the lattice vector  $\mathbf{RP}$  passes through a lattice point in plane 0.

Figure 20.6 illustrates some possible lattice shuffles for a  $q = 4$  twinning mode of a single lattice structure. No shuffles are required for the plane  $p = 2$ , but the atoms on the planes  $p = 1, 3$  undergo equal and opposite shuffles. A choice  $r = p$  and  $z_i = -w_i$  gives shuffles of  $\Delta^1 = -\frac{1}{2}\mathbf{w} + 4d\mathbf{m} = -\frac{1}{2}w_i\mathbf{b}_i$  for  $p = 1$  and an equal and opposite vector for  $p = 3$ . Note that these displacements are identical for any  $c_i$ , and that any other choice of  $z_i$  will give a vector  $\Delta^1$  which represents a longer and more complex displacement between the two  $K_1$  planes. The vectors  $\pm\frac{1}{2}w_i\mathbf{b}_i$  are formed by the shear from the vectors  $\pm(1/2)\mathbf{w}$  and thus are parallel to  $\pm\eta_2'$  in Fig. 2.4. (If the shuffles are regarded as preceding the shear, the choice  $r = p$  produces displacements  $\pm(1/2)\mathbf{w}$ , i.e. parallel to  $\pm\eta_2$ .)

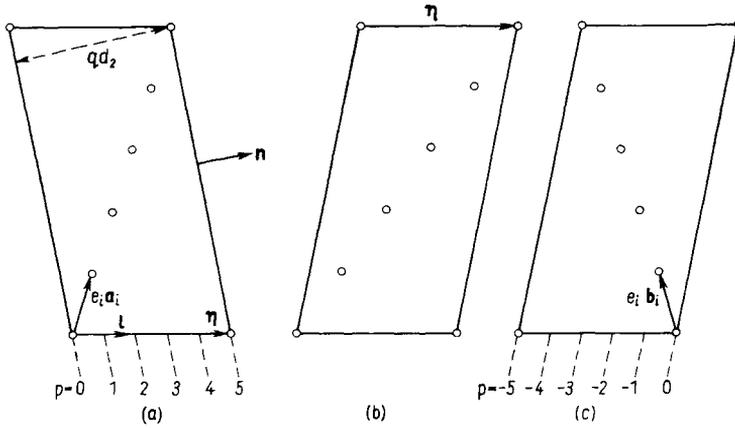


FIG. 20.6. Lattice sites in type II twinning for  $q=5$ .

The alternative choice  $r = q - p$  means that the shuffles are within the  $K_1$  plane, but not parallel to  $\eta_1$ . For  $p=1$ ,  $\Delta^I = \frac{1}{2}(4\mathbf{c} - \mathbf{w})$  if  $z_i = -w_i$ , and this choice of  $z_i$  thus cannot be generally valid as the shuffle would then vary with  $c_i$ . In order to minimize  $|\Delta^I|$ , the shuffle must relate the parent site to the nearest twin site in the same  $K_1$  plane. If this corresponds to a particular vector  $2\mathbf{c}_1 - \frac{1}{2}\mathbf{w}$  then, for any parent site specified by  $\mathbf{c} = \mathbf{c}_1 + \mathbf{f}$  (where  $f_i m_i = 0$ ), the correct shuffle is obtained by a choice  $z_i = \pm(2f_i + w_i)$  for  $p = 3, 1$  respectively.

Consider now the atomic movements during the formation of the twin; for  $q$  or  $\bar{q} > 2$ , these may have smaller magnitudes than the shuffles of the Bravais lattice points. The positions of the atoms in the parent are given by  $(u_i + \xi_{n,i})\mathbf{a}_i$  [see eqn. (5.8)], where  $\xi_{n,i}\mathbf{a}_i$  is the position of the  $n$ th atom in the unit cell relative to some chosen origin. The twin lattice will correspondingly have sites at  $(v_i + \xi_{n,i})\mathbf{b}_i$ . The unit cell origins may always be chosen so that  $\sum_n \xi_{n,i}\mathbf{a}_i = \sum_n \xi_{n,i}\mathbf{b}_i = 0$ ; if there are only two atoms in the primitive unit cell, this means that the origin is at the centre of symmetry.

After the lattice shear, the lattice sites are given by eqn. (85.15) and, for a rigid motif unit, the atom sites are given by the same expression with the addition of  $\xi_{n,i}\mathbf{a}_i$ . For orientation I, the twin lattice sites are given by eqn. (85.17) with  $\mathbf{b}_i = \mathbf{b}_i^{(I)}$  [eqn. (85.4)] whilst, for orientation IV, the twin sites are given by the negative of eqn. (85.17) with  $\mathbf{b}_i = \mathbf{b}_i^{(IV)}$ . As already noted, these two expressions are equivalent, and the twin lattice sites for orientations I and IV are identical. The atom sites in the twin are obtained by adding  $\xi_{m,i}\mathbf{b}_i^{(I)}$  or  $\xi_{m,i}\mathbf{b}_i^{(IV)}$  respectively to the vector representing the lattice sites. The sheared parent atom sites may thus be subtracted from the twin atom sites to obtain the atomic shuffle vectors  $\square^I$ . These vectors are given by

$$\square^I = \Delta^I - (\xi_{n,i} \pm \xi_{m,i})\mathbf{a}_i \pm 2\xi_{m,i}m_i\mathbf{m}. \tag{85.19}$$

The vector  $\square^I$  represents the displacement added to the homogeneous shear needed to transfer the  $n$ th atom of the parent unit cell to the  $m$ th atom site of the twin unit cell.

In order to specify it, the identifying subscripts  $n$  and  $m$  must be selected, in addition to the parameters defining  $\Delta^I$ . The alternative negative and positive signs are applicable to orientations I and IV respectively.

Now consider twinning of the second kind, in which the  $\boldsymbol{\eta}_I$  direction, but not necessarily the  $K_1$  plane, is rational. Because  $(u_i + v_i)\mathbf{a}_i$  is parallel to  $\mathbf{I}$  for orientation II [eqn. (85.6b)], it is clear that this vector is equal to the sum of the projections of the vectors  $u_i\mathbf{a}_i$  and  $v_i\mathbf{a}_i$  on  $\mathbf{I}$  (see Fig. 20.4). Hence eqn. (85.6b) may be written

$$su_i m_i \mathbf{I} = -(u_i - v_i)(\mathbf{a}_i \cdot \mathbf{I})\mathbf{I}. \tag{85.20}$$

As eqn. (85.7) is valid for orientation II, so also is eqn. (85.9) and, using this, eqn. (85.20) becomes

$$s(u_i - v_i)m_i \mathbf{I} = -2(u_i - v_i)(\mathbf{a}_i \cdot \mathbf{I})\mathbf{I}, \tag{85.21}$$

or, using eqn. (85.4),

$$(u_i - v_i)\mathbf{a}_i + s(u_i - v_i)m_i \mathbf{I} = -(u_i - v_i)\mathbf{b}_i. \tag{85.22}$$

This equation shows that the rational lattice vector  $(u_i - v_i)\mathbf{a}_i$  is sheared into its own twin  $-(u_i - v_i)\mathbf{b}_i$ . This vector is not a unique vector, but it follows from eqn. (85.21) that

$$(u_i - v_i)[sm_i + 2(\mathbf{a}_i \cdot \mathbf{I})] = 0. \tag{85.23}$$

The vector  $(u_i - v_i)\mathbf{a}_i$  is thus confined to a plane with unit normal  $\mathbf{n} = n_i \mathbf{a}_i^*$ , where  $Cn_i = sm_i + 2(\mathbf{a}_i \cdot \mathbf{I})$ . This rational plane is the conjugate twinning plane, or  $K_2$  plane, and all vectors in it are changed to their own twins by the lattice shear.

To discuss lattice shuffles with type II twinning, Bilby and Crocker again assumed that all lattice sites in any  $K_1$  plane move in the same way but, as  $K_1$  is irrational, this only includes the points lying along a single row parallel to  $\boldsymbol{\eta}_I$ . However, as all the sites on the  $K_2$  plane through the origin are sheared directly to their twin positions, so also are those on any other  $K_2$  plane which passes through lattice sites lying along a vector in the  $\mathbf{I}$  (i.e.  $\boldsymbol{\eta}_I$ ) direction from the origin. Shuffles may be necessary for lattice sites which do not lie on  $K_2$  planes of this type: let  $\boldsymbol{\eta}$  be a primitive lattice vector in the  $\boldsymbol{\eta}_I$  direction and suppose that it traverses  $\bar{q}$  lattice planes of type  $K_2$ . Because the configuration will be repeated as the interface moves through every  $\bar{q}$  planes of type  $K_2$ , only the lattice points in one such group need be considered. This is equivalent, for type II twinning, to the assumption that all lattice sites in any  $K_2$  plane move in the same way. The projection of  $\boldsymbol{\eta}$  along the unit normal  $\mathbf{n}$  is

$$\boldsymbol{\eta} \cdot \mathbf{n} = \bar{q}d_2, \tag{85.24}$$

where  $d_2$  is the spacing of the  $K_2$  planes. Now let any lattice point on the  $K_2$  plane a distance  $d_2$  from the origin be represented by the vector  $e_i \mathbf{a}_i$  (i.e.  $e_i n_i = d_2$ ), so that all the lattice sites on the intermediate group of  $(\bar{q} - 1)$  planes may be represented by  $\bar{p}e_i \mathbf{a}_i$ , where  $\bar{p}$  is a positive integer  $< \bar{q}$ .

The parent site at  $\bar{p}e_i \mathbf{a}_i$  will shear to a new position

$$\bar{p}e_i \mathbf{a}_i + \bar{p}sc_i m_i \mathbf{I}. \tag{85.25}$$

Let  $\bar{z} = \bar{z}_i \mathbf{a}_i$  define any lattice site of the parent structure in the  $K_2$  plane through the origin, so that  $\bar{z}_i n_i = 0$ . Equation (85.22) then shows that this site is sheared into a site  $-\bar{z}_i \mathbf{b}_i$  of the twin lattice. Relative to this twin site, the set of vectors  $\bar{r} e_i \mathbf{b}_i$ , where  $\bar{r}$  is a positive integer  $< \bar{q}$ , specifies all the lattice sites in a group containing  $\bar{q}$  of the  $K_2$  planes of the twin.<sup>†</sup> Hence the general expression for the twin sites is  $-(\bar{z}_i + \bar{r} e_i) \mathbf{b}_i$  and, substituting for  $\mathbf{b}_i^{(11)}$  from eqn. (85.4), this becomes

$$(\bar{z}_i - \bar{r} e_i) \mathbf{a}_i + 2\bar{r}(\mathbf{e} \cdot \mathbf{l}) \mathbf{l} - 2(\bar{\mathbf{z}} \cdot \mathbf{l}) \mathbf{l} \quad (85.26)$$

The relations of the lattice sites, sheared lattice sites and twin lattice sites are shown in Fig. 20.6(b) for  $\bar{q} = 5$ .

The vector defining the lattice shuffles is now obtained by subtracting eqn. (85.25) from eqn. (85.26) as

$$\Delta^{II} = \{\bar{z}_i - (\bar{p} + \bar{r}) e_i\} \mathbf{a}_i + 2\{\bar{r}(\mathbf{e} \cdot \mathbf{l}) - (\mathbf{z} \cdot \mathbf{l}) - \bar{p} s e_i m_i\} \mathbf{l} \quad (85.27)$$

This may be written in an alternative form by noting that the endpoint of the vector  $\bar{q} \mathbf{e} = \bar{q} e_i \mathbf{a}_i$  undergoes a vector displacement of  $\bar{q} s e_i m_i \mathbf{l}$  as a result of the lattice shear. From Fig. 20.6(b), it is clear that the projection of  $\bar{q} e_i \mathbf{a}_i$  on  $\mathbf{l}$  is equal to  $\boldsymbol{\eta}$  minus one-half of the displacement of the endpoint of  $\bar{q} e_i \mathbf{a}_i$ , so that

$$\bar{q}(\mathbf{e} \cdot \mathbf{l}) \mathbf{l} = \boldsymbol{\eta} - \frac{1}{2} \bar{q} s e_i m_i \mathbf{l} \quad (85.28)$$

Substituting into eqn. (85.27) now gives

$$\Delta^{II} = \{\bar{z}_i - (\bar{p} + \bar{r}) e_i\} \mathbf{a}_i - \{(\bar{p} + \bar{r}) s e_i m_i + 2(\mathbf{z} \cdot \mathbf{l}) \mathbf{l} + 2(\bar{r}/\bar{q}) \boldsymbol{\eta}\} \quad (85.29)$$

In considering the shuffles of a given lattice point in type II twinning,  $\bar{z}_i$  and  $\bar{r}$  are disposable parameters corresponding to  $z_i$  and  $r$  for type I twinning. In general  $\Delta^{II} = 0$  for all points only when  $\bar{q} = 1$  or 2. However, for any even value of  $\bar{q}$ , the lattice points on the plane  $\bar{p} = (1/2)\bar{q}$  need not shuffle, as it is always possible to choose  $\bar{r} = \bar{p} = \frac{1}{2}\bar{q}$  and  $\bar{z}_i = \bar{q} e_i - \boldsymbol{\eta}_i$ . If  $\bar{r} = \bar{q} - \bar{p}$  for all the lattice points, eqn. (85.28) or (85.29) shows with the above choice of  $\bar{z}_i$  that the shuffle vector is always parallel to  $\boldsymbol{\eta}_1$ . This may be compared with the condition  $r = p$  for type I twinning which gives a lattice shuffle vector parallel to  $\boldsymbol{\eta}'_2$ . For  $\bar{q} = 4$ , the shuffle vectors are  $\pm \frac{1}{2} \boldsymbol{\eta}$  but, as in the analogous case of type I twinning with  $q > 4$ , some of the shuffles described by this assumption look improbable for  $\bar{q} > 4$ . The alternative choice  $\bar{r} = \bar{p}$  gives shuffle displacements which are within the final position ( $K'_2$ ) of the conjugate twinning plane, but are not in general parallel to the conjugate twinning direction; these displacements are, of course, in  $K_2$  if the shuffles are supposed to precede the shear. The origin  $z_i$  for the twin sites now has to be chosen with regard to  $\mathbf{e}$ , in exact analogy with the type I shuffles in the  $K_1$  plane. Thus if  $\bar{z}_i = \bar{q} e_i - \boldsymbol{\eta}_i$ ,  $\Delta^{II} = (\bar{q} - 2\bar{p})(\mathbf{e} + s e_i m_i \mathbf{l} - \boldsymbol{\eta}/\bar{q})$ , and for  $\bar{q} = 4$  and  $\bar{p} = 1, 3$  the shuffles on the  $K'_2$  plane (following the shear) are  $\pm (2\mathbf{e} - \frac{1}{2} \boldsymbol{\eta} + 2s e_i m_i \mathbf{l})$ . The minimum shuffle displacement of this

<sup>†</sup>A different specification of the type II twin sites was used in the first edition; the present formalism follows that of Bilby and Crocker (1965).

type will have this form for some particular value  $\mathbf{e}_1$  of  $\mathbf{e}$  and, for other representative vectors  $\mathbf{e}$ , a different choice of  $\bar{z}_i$  must be made to give the same minimum shuffle. The condition for this is most readily seen by noting that the shuffle displacement in  $K_2$  (preceding the shear) is  $2\mathbf{e}_1 - \frac{1}{2}\boldsymbol{\eta}$  and, writing  $\mathbf{e} = \mathbf{e}_1 + \mathbf{f}$ , it follows that  $\mathbf{z} = z_i\mathbf{a}_i$  should be chosen as  $4\mathbf{e}_1 + 2\mathbf{f} - \boldsymbol{\eta}$ . Figure 20.6 shows the lattice sites for the parent, sheared parent and type II twin for  $\bar{q} = 5$ .

The structure shuffles may now be defined in the same way as those for type I twinning. The general expression is

$$\square^{II} = \Delta^{II} - (\xi_{n,i} \pm \xi_{m,i})\mathbf{a}_i \pm (\xi_m \cdot \mathbf{l}) \quad (85.30)$$

where the alternative positive and negative signs relate to orientations II and III respectively. The vector  $\square^{II}$  represents the displacement added to the homogeneous shear which is needed to transfer the  $n$ th atom of the parent unit cell to the  $m$ th site of the twin cell.

Equations (85.19) and (85.30) are the general expressions for the atom shuffles in any kind of twin formation, and some properties of these expressions in particular cases of interest will now be discussed. All shuffles are zero in simple lattice structures when  $q$  or  $\bar{q} = 1$  or 2, and in predicting the twinning elements of these structures the lowest shear modes for  $q$  or  $\bar{q} = 2$  should thus be considered first. Higher values of  $q$  or  $\bar{q}$ , with associated shuffles, need be considered only if modes are observed experimentally which do not correspond to  $q$  or  $\bar{q} = 2$  predictions. This is the procedure used by Jaswon and Dove, and described above; as already noted, it gives excellent agreement with experimental results.

Now consider double lattice structures, which cover most other pure metals known to form mechanical twins. Because these structures are centrosymmetric, there are only two possible orientation relations, and the shuffles obtained by choosing positive or negative signs in eqns. (85.19) or (85.30) must thus be equivalent. It is convenient to write  $\xi_{1,i} = \xi_i$  and  $\xi_{2,i} = -\xi_i$  for such structures. Then if  ${}^l(n, m)$  represents the shuffle associated with the movement of the  $n$ th atom of the parent to the  $m$ th site of the twin, eqn. (85.19) shows that  ${}^I(1, 1)$  for orientation I is equal to  ${}^I(1, 2)$  for orientation IV. There are similar relations of this kind for all the double lattice shuffles.

Suppose first that the twinning elements in double lattice structures are such that  $q$  or  $\bar{q} = 1$  or 2. Then the lattice shuffles are zero, and there are four possible types of structure shuffle, which are respectively

$$\begin{aligned} |\square^I(1, 1)| &= |\square^I(2, 2)| = 2\xi_i m_i, \\ |\square^I(1, 2)| &= |\square^I(2, 1)| = 2[|\xi|^2 - (\xi_i m_i)^2]^{1/2}, \\ |\square^{II}(1, 1)| &= |\square^{II}(2, 2)| = 2\xi \cdot \mathbf{l}, \\ |\square^{II}(1, 2)| &= |\square^{II}(2, 1)| = 2[|\xi|^2 - (\xi \cdot \mathbf{l})^2]^{1/2} \end{aligned} \quad (85.31)$$

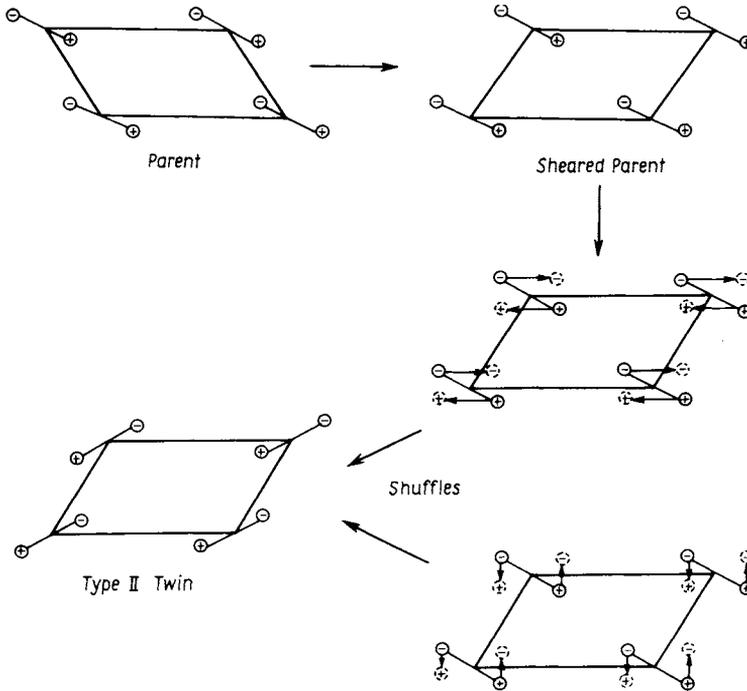


FIG. 20.7. Possible shuffles for a double lattice structure type II twin with  $\bar{q}=1$  or 2. Circles marked + or - represent atoms at equal distances above and below the plane of projection. The sheared parent shows the structure after applying a homogeneous shear to the lattice points, and the two shuffle mechanisms may be compared to those in Fig. 20.2. In one mechanism, the atoms shuffle parallel to  $\eta_1$ ; in the other they move normal to  $\eta_1$  but not normal to  $K_1$ .

The first two of these mechanisms, applicable to a type I twin, were described as the *Y* and *X'* mechanisms respectively earlier in this chapter, and are shown in Fig. 20.2. The second two mechanisms are correspondingly shown in Fig. 20.7.

For given twinning elements, a possible hypothesis is that the operative shuffle mechanism will be that for which  $|\square|$  is smallest; this could result in either a type I or a type II orientation relation. It follows from eqn. (85.31) that, in particular cases, double lattice structures may be able to twin without atomic shuffles. Thus if  $\xi \cdot m = 0$ , the *Y* mechanism reduces to zero shuffles, whilst if  $\xi \cdot l = 0$ , the first of the type II shuffles similarly disappears. These two possibilities arise if the motif unit may be chosen respectively in the  $K_1$  plane or in the plane normal to the  $\eta_1$  direction, and it may be readily seen that they apply also to type I and type II twinning in centrosymmetric structures where there are more than two atoms in the motif units. On p. 863, however, it was pointed out that the division into shear plus shuffles may be made in different ways. If the motif unit is not regarded as rigid, but each atom is instead displaced individually by the twinning shear, the condition for no shuffles in type I twinning will be unchanged, but

the condition for absence of shuffles in type II twinning then becomes a motif unit lying in the  $K_2$  plane rather than in the plane normal to  $\eta_1$ .

The motif unit is not uniquely defined, even for a double lattice structure, as in principle it may link an atom of one lattice to any atom of the interpenetrating lattice. However, the analysis of this section, in which the motif units are treated as rigid during the shear, implies that the separation of the atoms in a unit must be small. The analysis is purely a matter of mathematical convenience, and the only physical reality is the net displacements of the individual atoms, but it is clear that the above treatment corresponds to the natural assumption that each atom goes to the nearest available twin site. Thus it is possible to impose a restriction that  $|\xi_{n,i}| \leq \frac{1}{2}$ , so that unsuitable large motif units are excluded. This assumption is not adequate for large unit cells, especially if it allows atomic interchange shuffles (see p. 885), as in the formal theory of twinning of superlattices, and it is probably better then to use an alternative division into shear plus shuffles (Christian and Laughlin, 1988). It should also be noted that even the above restriction does not necessarily define the unit uniquely but, if the plane of shear is rational, the most plausible motif unit can usually be found by inspection.

In eqns. (85.31), the vector sum of the atom shuffles is zero in any mechanism, because  $\square^1(1, 1) = -\square^1(2, 2)$ , etc. From eqns. (85.19) and (85.30), we see that the vector sum of all the atom shuffles within the unit cell defined by the value of  $q$  or  $\bar{q}$  is zero for double lattice structures, provided we can ensure that the vector sum of the lattice shuffles  $\sum \Delta^I$  or  $\sum \Delta^{II}$  is zero. The general condition for this is readily obtained. Let  $\Delta^I(p, r)$  be the type I twin lattice shuffle associated with a lattice point on the  $p$ th  $K_1$  plane of the parent and the  $r$ th  $K_1$  plane of the twin, and  $\Delta^{II}(\bar{p}, \bar{r})$  be the corresponding lattice shuffle defined with respect to  $K_2$  planes in type I twinning. Then from eqn. (85.18), if the two sets of values of  $z_i$  which specify the shuffles  $\Delta^I(p, r)$  and  $\Delta^I(q-p, q-r)$  are chosen so that

$$z_i(p, r) + z_i(q-p, q-r) = -2w_i \tag{85.32}$$

it follows that

$$\Delta^I(p, r) = -\Delta^I(q-p, q-r)$$

and

$$\sum \Delta^I = 0$$

Similarly for type II twinning, the two values of  $\bar{z}_i$  corresponding to  $(\bar{p}, \bar{r})$  and  $(\bar{q} - \bar{p}, \bar{q} - \bar{r})$  shuffles respectively may be chosen so that

$$\bar{z}_i(\bar{p}, \bar{r}) + \bar{z}_i(\bar{q} - \bar{p}, -\bar{q} - \bar{r}) = 2(\bar{q}e_i - \eta_i) \tag{85.33}$$

and this gives

$$\Delta^{II}(\bar{p}, \bar{r}) = -\Delta^{II}(\bar{q} - \bar{p}, -\bar{q} + \bar{r})$$

or

$$\sum \Delta^{II} = 0.$$

Provided that pairs of values of  $z_i$  or  $\bar{z}_i$  are always chosen in this way, the net lattice shuffles are zero, and so also are the net structure shuffles in double lattice structures. For  $q = 1$  or  $2$ , the result is trivial, as there are no lattice shuffles, but it is now possible to discuss the important case of  $q = 4$ .

The restrictions on the motif unit ( $|\xi_{n,i}| \leq \frac{1}{2}$ ) ensure that this unit is not disrupted in the  $q = 1, 2$  shuffles already discussed. For  $q = 4$ , the shuffles for sites on the planes defined by  $p = 2, 4$  will correspond to those already discussed, as the lattice sites are sheared directly to their twin positions. For the shuffles on  $p = 1$  and  $p = 3$ , assume first that the lattice shuffle vectors  $\Delta^1$  are parallel to  $K_1$  (see p. 871). From eqns. (85.18) and (85.19), with the restriction  $p + r = q$ , this gives for the  $p = 1$  shuffles

$$\begin{aligned}\square^1(1, 1) &= (2c_i + \frac{1}{2}w_i + z_i)\mathbf{a}_i - 2\xi_i m_i \mathbf{m}, \\ \square^1(2, 2) &= (2c_i + \frac{1}{2}w_i + z_i)\mathbf{a}_i + 2\xi_i m_i \mathbf{m}, \\ \square^1(1, 2) &= (2c_i + \frac{1}{2}w_i + z_i)\mathbf{a}_i - 2\xi_i + 2\xi_i m_i \mathbf{m}, \\ \square^1(2, 1) &= (2c_i + \frac{1}{2}w_i + z_i)\mathbf{a}_i + 2\xi_i - 2\xi_i m_i \mathbf{m}.\end{aligned}\tag{85.34}$$

By choosing  $z_i = -w_i$  for  $\square^1(1, 2)$  and  $z_i = -4c_i$  for  $\square^1(2, 1)$ , a shuffle mechanism is obtained in which

$$\square^1(1, 2) = -\square^1(2, 1) = (2c_i - \frac{1}{2}w_i)\mathbf{a}_i - 2\xi_i + 2\xi_i m_i \mathbf{m}.\tag{85.35}$$

This means that the vector sum of the shuffles associated with one lattice site of the parent is zero. This can only be achieved by using different  $z_i$  values for the  $\square^1(1, 2)$  and  $\square^1(2, 1)$  shuffles; that is, the two atoms around one parent lattice site move to different twin lattice sites, and the motif unit is disrupted. The equations for the  $p = 3$  plane are equal to eqn. (85.34) with  $2c_i$  replaced by  $-2c_i$  and  $\frac{1}{2}w_i$  replaced by  $3w_i/2$ . The same shuffle mechanism is obtained by choosing  $z_i = 4c_i - 2w_i$  for the  $\square^1(2, 1)$  shuffle and  $z_i = -w_i$  for the  $\square^1(1, 2)$  shuffle.

The lattice shuffle vectors  $\Delta^1$  are not completely defined when the motif units are disrupted, as the displacement of the lattice point may be associated with that of either atom. However, if the  $\xi_1$  atom is chosen for the site on the plane  $p = 1$ , the  $\xi_2$  atom must be chosen for the site on the plane  $p = 3$ , and vice versa, and the two values of  $z_i$  then satisfy eqn. (85.32). This is necessary because  $\sum \Delta^1$  must clearly be zero if the sum of  $\square^1$  for each separate lattice point is zero. As noted above, the parameters for any double lattice structure can always be chosen so that  $\sum \square^1 = \sum \Delta^1 = 0$  over the whole unit cell, whatever the value of  $q$ . When  $q$  is greater than four, however, it is not possible to make the sum of the individual shuffles at any lattice site equal to zero. This arises when  $q = 4$  only because each sheared parent lattice site on the planes  $p = 1, 3$  is midway between two twin lattice sites. For this reason, the  $q = 4$  shuffles are relatively simple, and may be important in practice.

The same values of  $z_i$  for the  $\square^1(1, 1)$  and  $\square^1(2, 2)$  shuffles respectively lead to an alternative shuffle mechanism in which

$$\square^1(1, 1) = -\square^1(2, 2) = (2c_i - \frac{1}{2}w_i)\mathbf{a}_i - 2\xi_i m_i \mathbf{m}.\tag{85.36}$$

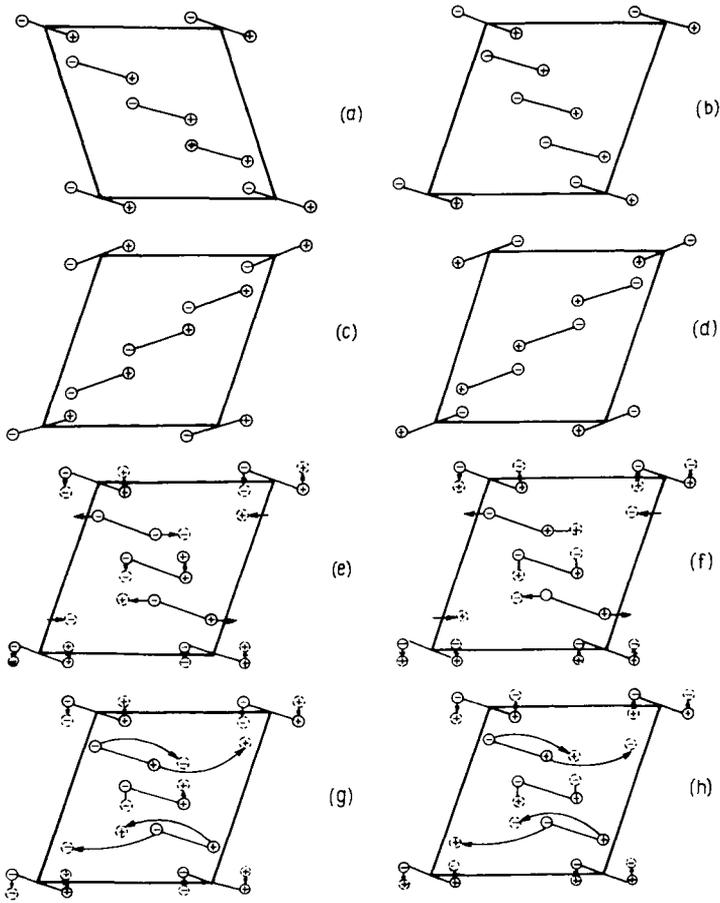


FIG. 20.8. Possible shuffling mechanisms in double lattice structure twins with  $q=4$ : (a) parent structure; (b) sheared parent; (c) type I twin; (d) type II twin; (e) shuffle given by eqn. (85.37) ( $X$  mechanism); (f) type II shuffle given by eqn. (85.37); (g) type I shuffle given by eqn. (85.36); (h) type II shuffle given by eqn. (85.38).

The mechanism (85.36) involves atom movements which are generally large and not parallel to the  $K_1$  plane, in contrast to mechanism (85.35) where the movements are parallel to  $K_1$  and may be quite small. Both mechanisms are illustrated in Fig. 20.8; that given by (85.35) is the  $X$  mechanism already described and also shown in Fig. 20.3.

The atomic shuffles for type II twins in which  $\bar{q}=4$  may be investigated in a similar way. Using eqns. (85.29) and (85.30), with  $\bar{\varepsilon}_i = 2e_i - \eta_i$  for  $\square^{II}(1, 1)$  and  $\square^{II}(1, 2)$  and  $\bar{\varepsilon}_i = 0$  for  $\square^{II}(2, 2)$  and  $\square^{II}(2, 1)$ , we find that the two shuffle mechanisms for the  $\bar{p}=1$  plane are

$$\square^{II}(1, 1) = -\square^{II}(2, 2) = -\frac{1}{2}\eta + 2(\xi \cdot \mathbf{l})\mathbf{l}, \tag{85.37}$$

and

$$\square^{\text{II}}(1, 2) = -\square^{\text{II}}(2, 1) = -\frac{1}{2} \boldsymbol{\eta} + 2\xi + 2(\xi \cdot \mathbf{l})\mathbf{l}. \quad (85.38)$$

The same two mechanisms are obtained for the plane  $\bar{\rho} = 3$  with the choices  $\bar{z}_i = \bar{q}_i - \boldsymbol{\eta}$  for  $\square^{\text{II}}(2, 1)$  and  $\square^{\text{II}}(2, 2)$  and  $\bar{z}_i = 2\bar{q}_i - 2\boldsymbol{\eta}$ , for  $\square^{\text{II}}(1, 1)$  and  $\square^{\text{II}}(1, 2)$ . These values of  $\bar{z}_i$  satisfy eqn. (85.33) to ensure that  $\sum \Delta^{\text{II}} = 0$ , provided that, as in the type I shuffles, different atoms of a motif pair are used to define the motion of the lattice points on  $\bar{\rho} = 1$  and 3. The two shuffle mechanisms are also shown in Fig. 20.8; that given by eqn. (85.37) represents atom movements which are parallel to  $\boldsymbol{\eta}_1$  and may be quite small.

Now consider the problem of compound twins. A given set of twinning elements can give a type I twin if  $K_1$  and  $\boldsymbol{\eta}_2$  are rational and a type II twin if  $K_2$  and  $\boldsymbol{\eta}_1$  are rational. When all four elements are rational, we have a compound twin, and there is the possibility of two different orientation relations existing for one set of twinning elements (four if the structure is not centrosymmetric). There will then correspondingly be two different shuffle mechanisms, and the preferred orientation may be decided by the easier shuffle mechanism. However, the conditions for the formation of compound twins are rather restrictive and, in all crystal systems except rhombohedral and cubic,<sup>†</sup> the plane of shear must be a plane of symmetry (Crocker, 1959). In fact, the plane of shear is found to be a mirror plane in most of the operative modes of rhombohedral and cubic crystals also. Hence it follows (see p. 867) that in most single lattice structures all four orientation relations are identical in compound twins. For double lattice structures, orientations I and II will not be equivalent unless the plane of shear is a symmetry plane of the atomic arrangement, i.e. unless the motif unit may be chosen to lie in or normal to the plane of shear. In other cases of compound twins in double lattice structures, there will be two orientation relations to consider. This may be seen in Fig. 20.8, where the distinction between (e) and (f) and between (g) and (h) disappears if the motif unit is in, or normal to, the plane of shear.

Now consider briefly the phenomenon of double twinning, mentioned on p. 867. Consider the application of two successive twinning shears to a region of crystal. If a macroscopic twin is produced by the first deformation, the retwinning represented by the second shear will not generally be possible unless there is considerable additional deformation in either the parent or the primary twin. Assuming the re-orientated region to be plate-shaped, its macroscopic habit plane will be determined by the  $K_1$  plane of the first twin and by the amount of additional deformation which has occurred in the parent. If such situations arise, they are clearly of interest in producing a set of pseudotwinning elements, but they introduce no new principles in the crystallography of twinning. A different type of double twinning, however, results from the assumption that two twinning shears are applied simultaneously rather than successively to a region of parent. This might arise physically, for example, if a small twin nucleus retwins, and the re-orientated nucleus then grows into a macroscopic twin by the combined action of the two shears. The final product may then possess twin elements different from those of the

<sup>†</sup>Compound twins cannot be formed in triclinic crystals.

constituent twins, and in a sense the resolution into components has only formal significance as far as the macroscopic growth is concerned. A discussion of both these types of double twinning was given by Crocker (1962), following suggestions by Couling *et al.* (1959) and by Reed-Hill (1960) that certain anomalous twinning modes observed in magnesium are in fact due to double twinning.

From the more general result proved in Section 9 for invariant plane strains, it follows that the resultant of two simple shears will not itself be a simple shear unless either the  $K_1$  plane or the  $\eta_1$  direction is common to the two components. This is possible in a formal sense in cubic structures, for example, but the net result is to restore the original parent orientation, so that a simple combination of two twinning shears does not give physically significant results. If the above condition on the  $K_1$  plane or  $\eta_1$  direction is relaxed, the resultant deformation will not contain an invariant plane, but may contain an undistorted plane. There is thus the additional possibility of combining two twinning shears and a pure rotation to give an equivalent simple shear. In analysing this type of double twinning operation, Crocker points out the close similarity to the theory of martensite crystallography.

The two component shears and the resultant deformation must all have a principal strain equal to zero (p. 59), and this restricts the combinations to be considered to those in which the two twinning directions and the normals to the two  $K_1$  planes are all coplanar. The two planes of shear are thus coincident, and the problem is essentially two-dimensional. Consideration of the various possible combinations of type I and type II twins then shows that the plane of shear must be rational and, furthermore, all eight twinning elements of the two component twinning modes must also be rational, so that only compound modes may be combined in this way.

The combined effect of the three component deformations is to produce an equivalent simple shear with the same rational plane of shear, so that this also represents a compound twinning mode if it satisfies one of the usual types of twin orientation relation. The general solution, however, is one in which all four elements of the equivalent shear mode are irrational, and this is thus an example of the more general type of twin which has to be defined in the way indicated on p. 860. The possible existence of "twins" with four irrational twinning elements is a remarkable result of the theory of double twinning. The assumptions of this theory nevertheless appear to be rather artificial and it seems preferable to generalize the classical theory of deformation twinning by beginning directly with the proposed definition (Bilby and Crocker, 1965) that a twinning shear is any shear which restores the lattice or a superlattice in a new orientation. The relevant theory for lattices (i.e. excluding detailed consideration of atomic shuffles) was first given by Bevis and Crocker (1968, 1969) and we follow their treatment but use matrix rather than tensor notation (Christian, 1970).

The theory is developed in terms of certain properties of the correspondence matrix  $C$  defined in eqn. (8.4). In the case of twinning, the matrices  $C$ ,  $S$  and  $L$  are all unimodular and  $S$  has the form

$$S = I + s\mathbf{lm}' \quad (85.39)$$

in which (as before)  $\mathbf{l}$  is a unit vector in the  $\eta_1$  direction and  $\mathbf{m}$  the unit normal to the  $K_1$  plane so that, in a general coordinate system with metric  $\mathbf{G}$  (see p. 34),  $\mathbf{l}'\mathbf{G}\mathbf{l} = \mathbf{m}'\mathbf{G}^{-1}\mathbf{m} = 1$  and  $\mathbf{m}'\mathbf{l} = 0$ . It follows from the expression for the length of a vector  $|\mathbf{u}|^2 = \mathbf{u}'\mathbf{G}\mathbf{u}$  and from the equivalent expression after a rigid body rotation represented by the orthogonal matrix  $\mathbf{L}$  that  $\mathbf{L}'\mathbf{G}\mathbf{L} = \mathbf{G}$ , so that

$$\mathbf{X} = \mathbf{S}'\mathbf{G}\mathbf{S} - \mathbf{C}'\mathbf{G}\mathbf{C} = 0 \quad (85.40)$$

is a null matrix. Combining eqns. (85.39) and (85.40) gives

$$\mathbf{X} = \mathbf{G} + s\mathbf{G}\mathbf{l}\mathbf{m}' + s\mathbf{m}\mathbf{l}'\mathbf{G} + s^2\mathbf{m}\mathbf{m}' - \mathbf{C}'\mathbf{G}\mathbf{C} \quad (85.41)$$

An expression for the magnitude of the shear which depends only on the correspondence matrix may now be obtained by taking the trace of the matrix product  $\mathbf{X}\mathbf{G}^{-1}$  and equating this to zero, to give

$$s^2 = \text{tr}(\mathbf{C}'\mathbf{G}\mathbf{C}\mathbf{G}^{-1}) - 3 \quad (85.42)$$

An equivalent equation may be obtained by inverting eqn. (85.40) to form a matrix  $\mathbf{S}^{-1}\mathbf{G}^{-1}(\mathbf{S}^{-1})' - \mathbf{C}^{-1}\mathbf{G}^{-1}(\mathbf{C}^{-1})'$  which is also identically equal to zero. Combination of the resultant equation for  $s^2$  with eqn. (85.42) then leads to a restriction on the correspondence matrix

$$\text{tr}(\mathbf{C}'\mathbf{G}\mathbf{C}\mathbf{G}^{-1}) = \text{tr}\left\{(\mathbf{C} - \mathbf{1})'\mathbf{G}^{-1}\mathbf{C}^{-1}\mathbf{G}\right\} \quad (85.43)$$

Once the correspondence is defined, it fixes not only the shear but also  $\mathbf{l}$  and  $\mathbf{m}$ . Consider, for example, the scalar combination  $2m_1m_2X_{12} - m_1^2X_{22} - m_2^2X_{11}$  which is identically equal to zero from eqn. (85.40). When this expression is formed from eqn. (85.41), the terms in  $s$  and  $l_1$  all vanish and therefore

$$Y_{11}m_2^2 - 2Y_{12}m_1m_2 + Y_{22}m_1^2 = 0 \quad (85.44)$$

where

$$\mathbf{Y} = \mathbf{G} - \mathbf{C}'\mathbf{G}\mathbf{C} \quad (85.45)$$

Equation (85.44) is a quadratic in the ratio  $m_2/m_1$  and there are two similar equations for the ratios  $m_3/m_2$  and  $m_1/m_3$ ; if  $\mathbf{C}$  is known, the three equations give two possible solutions for the components of  $\mathbf{m}$ . The three components of  $\mathbf{l}$  may be determined directly from  $\mathbf{m}$  and  $s$ , or may be derived from three similar quadratic equations of the form

$$Z_{11}l_2^2 - 2Z_{12}l_1l_2 + Z_{22}l_1^2 = 0 \quad (85.46)$$

where

$$\mathbf{Z} = \mathbf{G}^{-1} - (\mathbf{C}^{-1})'\mathbf{G}^{-1}\mathbf{C}^{-1} \quad (85.47)$$

Consider now the classical theory of deformation twinning. It follows from eqn. (85.4) that, for a type I twin, the matrix  $\mathbf{L}$  of eqn. (8.4) is given by

$$\mathbf{L} = -\mathbf{I} + 2\mathbf{G}^{-1}\mathbf{m}\mathbf{m}' \quad (85.48)$$

Also, eqn. (85.12) expressed in a general coordinate system is

$$s\mathbf{l} = 2[\mathbf{G}^{-1}\mathbf{m} - (\mathbf{g}/\mathbf{m}'\mathbf{g})] \quad (85.49)$$

so that eqn. (85.39) becomes

$$\mathbf{S} = \mathbf{I} + 2\mathbf{G}^{-1}\mathbf{m}\mathbf{m}' - 2\mathbf{g}\mathbf{m}'/(\mathbf{m}'\mathbf{g}) \quad (85.51)$$

The correspondence matrix for type I twinning is thus

$$\mathbf{C} = \mathbf{L}\mathbf{S} = -\mathbf{I} + 2\mathbf{g}\mathbf{m}'/(\mathbf{m}'\mathbf{g}) \quad (85.51)$$

and the form of the equation remains unchanged if we substitute any vector parallel to  $\boldsymbol{\eta}_2$  for  $\mathbf{g}$  and any reciprocal lattice vector normal to  $K_1$  for  $\mathbf{m}$ . A similar development gives the correspondence matrix for type II twinning as

$$\mathbf{C} = -\mathbf{I} + 2\mathbf{l}\mathbf{n}'/(\mathbf{n}'\mathbf{l}) \quad (85.52)$$

The two correspondence matrices just derived are independent of the metric tensors  $\mathbf{G}$  and  $\mathbf{G}^{-1}$  and so give rise to classical twins in all lattices; they have the property that  $\mathbf{C} = \mathbf{C}^{-1}$ , so that eqn. (85.43) is automatically satisfied. They also have the property that  $\text{tr } \mathbf{C} = -1$  and this is a useful necessary, although not sufficient, condition for  $\mathbf{C}$  to represent a deformation twin of classical type.

It remains to investigate the possibility of other correspondences which lead to non-classical twins. Following the Jaswon–Dove approach, it is useful to rewrite eqn. (85.42) as an inequality in order to list all chosen values of  $s_{\max}$ . In particular, this gives for the cubic system

$$\text{tr}(\mathbf{C}'\mathbf{C}) \leq s_{\max}^2 + 3 \quad (85.53)$$

As stated on p. 56, the columns of  $\mathbf{C}$  are the components of the vectors specifying the cell into which the reference cell defined by  $\mathbf{a}_i$  is deformed by  $\mathbf{S}$ . The lattice is not reproduced if any of the components of  $\mathbf{C}$  is irrational. When  $\mathbf{a}_i$  define a primitive unit cell, the point lattices of parent and twin are identical if the columns of  $\mathbf{C}$  represent lattice vectors. However, if the base vectors define a base-centred cell, it is additionally necessary that the sum of the first two columns of  $\mathbf{C}$  must be twice a lattice vector, for a body-centred cell the sum of all three columns of  $\mathbf{C}$  must be twice a lattice vector, and for a face-centred cell the sum of any two columns must be twice a lattice vector. When the elements of  $\mathbf{C}$  are rational but do not satisfy these conditions, the point lattice produced by  $\mathbf{S}$  differs from the parent lattice but has a superlattice in common, so that a twin may be produced by combining the shear with a shuffling of some fraction of the lattice sites. If some elements of  $\mathbf{C}$  are fractions, a matrix  $\mathbf{W} = m\mathbf{C}$  with only integral elements may be defined, and the inequality (85.53) becomes

$$\text{tr}(\mathbf{W}'\mathbf{W}) \leq m^2(S_{\max}^2 + 3) \quad (85.54)$$

Using a trial-and-error procedure, Bevis and Crocker list 10 correspondences for cubic systems in which  $m = 1$ ,  $s_{\max}^2 = 9$ , 19 correspondences for which  $m = 2$ ,  $s_{\max}^2 = 3.5$ , and 31 correspondences for which  $m = 4$ ,  $s_{\max}^2 = 2.5$ . For primitive lattices, the fraction of the

lattice points sheared direct to twin positions is  $1/m$ , but it may be  $2/m$ ,  $1/m$  or  $\frac{1}{2}m$  for centred lattices.

New unimodular matrices may be derived from any given  $C$  by interchanging rows or columns, or changing their signs. In the cubic system, these operations simply lead to equivalent variants of  $C$  but, in systems of lower symmetry, the numbers of non-equivalent variants which may be derived from each cubic correspondence are: tetragonal, nine; rhombohedral, 16; orthorhombic and hexagonal, 36; monoclinic, 144; triclinic, 576. However, many of these do not represent possible twins because the restriction of eqn. (85.43) involves lattice parameters in non-cubic systems.

Consider the set of planes and directions  $K_1$ ,  $K_2$ ,  $\eta_1$  and  $\eta_2$  which define a mode arising from a particular correspondence  $C$ . The associated conjugate mode is then obtained by interchanging  $K_1$  and  $K_2$  and  $\eta_1$  and  $\eta_2$  and also has the correspondence  $C$ . A different but closely related twinning mode is obtained by interchanging the indices of  $K_1$  and  $\eta_1$  and of  $K_2$  and  $\eta_2$  and reversing the sign of either the old or the new  $K_2$  and  $\eta_1$ . This new mode arises from the correspondence  $(C^{-1})'$ , i.e. the transpose of the inverse of the original correspondence, and the interchange leads strictly to an equivalent shear (of the same magnitude) in the reciprocal lattice provided that the direct and reciprocal metrics are also interchanged. However, the new shear mode is valid for all possible reciprocal lattice parameters and, as the direct and reciprocal lattices belong to the same crystal system, these parameters may be chosen to equal those of the original direct lattice. Thus the correspondences  $C$  and  $(C^{-1})'$  define separate but related shear modes for a given lattice. If the lattice is centred, the fraction of the lattice points sheared directly to twin positions will normally not be the same for the two modes.

For non-conventional twinning modes, the elements derived from  $C$  are different from those derived from  $C^{-1}$ . The apparent difference conceals a close relationship; the shear matrix derived from  $C^{-1}$  represents in the twin basis an equal and opposite shear to that derived in the parent basis from  $C$ . Thus the new twinning mode would convert the twin produced by the original mode back to the original parent orientation. It follows that the indices of the  $K_1$  plane in a non-conventional twinning mode are different when referred to parent and twin bases and are described by the correspondences  $C$  and  $C^{-1}$  respectively; the same applies to  $K_2$ ,  $\eta_1$  and  $\eta_2$ . Moreover, the relation between the twinning elements derived from  $C$  and  $(C^{-1})'$  applies equally to those derived from  $C^{-1}$  and  $C'$ .

The correspondence matrices may be divided into seven different classes by considering the various crystallographic degeneracies which arise from relations between  $C$ ,  $C^{-1}$ ,  $C'$  and  $(C^{-1})'$ . The most general class of twinning mode arises when none of the three derived matrices is crystallographically equivalent to  $C$  and the indices  $K_1$ ,  $K_2$ ,  $\eta_1$  and  $\eta_2$  associated with the mode and its inverse are then all different, so that eight sets of twinning elements are associated with the four correspondences. In the other classes, there are various predicted relations between the individual twinning elements; for example, if  $C$  and  $C^{-1}$  are crystallographically equivalent, but not identical, the twinning elements given by the two correspondence matrices may be written  $ABCD$  and  $BADC$  respectively, whereas if  $C$  and  $(C^{-1})'$  are crystallographically equivalent, the twinning elements obtained from  $C$  and  $C^{-1}$  have the forms  $ABAB$  and  $EFEF$  respectively. Both of these are degenerate cases of

non-conventional twinning modes, but two of the seven classes include the condition  $\mathbf{C} = \mathbf{C}^{-1}$  which gives rise to classical twinning modes. Of the 60 correspondence matrices mentioned above, 42 represent conventional or classic modes and 18 represent non-conventional modes.

To end this section, consider the problem of defining deformation twinning in alloys. Laves (1952) and many later authors have emphasized that, in the absence of shuffles, it is impossible to twin many superlattice structures as a homogeneous shear produces a new superlattice which can be represented as a different ordering of the atoms of the various chemical species on the sites of the disordered lattice. From one point of view, a superlattice is just a particular example of a multiple lattice structure, and no new principles need be invoked in deducing the possible twinning modes which will, in general, require structure shuffles in addition to the displacements of the simple shear. However, these shuffles would usually simply effect interchanges of some of the atoms, and are thus extremely improbable during deformation twinning. In discussing the possible twin modes of superlattices, it is thus necessary to distinguish "interchange" shuffles from the lattice and structure shuffles considered above (Christian and Laughlin, 1988). In fact, some superlattice structures have been found to undergo deformation "twinning" without the necessary shuffles in a mode appropriate to the disordered structure and, following Laves, this is often referred to as "pseudotwinning". The sheared region which comprises the pseudotwin generally has a lower symmetry than the parent structure; for example, if the usual b.c.c. twinning mode is applied to the B2 ( $L2_0$ ) superlattice, the pseudotwin has a structure with orthorhombic symmetry.

The change of structure implies that the new region will have a higher free energy per atom, and the change should strictly be regarded as a particular case of a stress-induced fully coherent martensitic transformation with no change of volume. This argument may be extended to any non-ideal solid solution, to which the application of a twinning shear produces a change in the pair correlations (Laves, 1966) or traps interstitials in different positions (Magee *et al.*, 1971). In such a solid solution, "true" twinning is thus strictly not possible, although the structural and symmetry changes may be too small in practice to be detected experimentally. The extreme form of this conclusion (Cahn, 1977) is that, as there are no absolutely pure metals or perfectly random solid solutions, deformation "twinning" will always in principle be accompanied by some structural and energetic changes. Cahn suggests that, because the assignment of some limit to the permissible structure change would be arbitrary, we should not attempt to distinguish between twinning and "twinning", i.e. between true and pseudotwins, but should recognize that changes of structure and symmetry are always involved in the process which we describe as deformation twinning. However, it is convenient to retain the term pseudotwinning for structures with nearly perfect long-range order because, as explained above, the operation of a twinning mode of the disordered system may then give a relatively large change in structure and symmetry.

As the normal twinning modes of the (disordered) f.c.c. and b.c.c. structures both have  $q=4$  for simple cubic lattices and  $q=8$  for their reciprocal lattices (b.c.c. and f.c.c. respectively), it follows that lattice (interchange) shuffles will only be avoided for a cubic

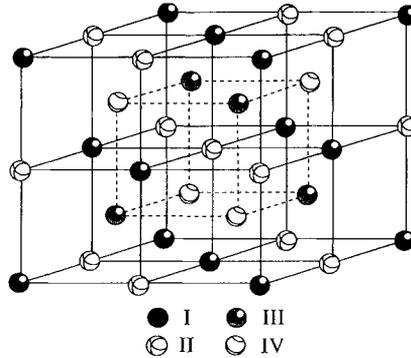


FIG. 20.9. Superlattice structures derived from b.c.c. structure (after Christian and Laughlin, 1987). The basic superlattice structures have site occupancy as follows: B2, *B* atoms occupy sites I and II and *A* atoms occupy sites III and IV; B32, *B* atoms occupy sites I and II and *A* atoms occupy sites I and IV; DO<sub>3</sub>, *B* atoms occupy sites I and *A* atoms occupy sites II, III and IV.

superlattice which has the same Bravais lattice as the disordered structure. This is only possible for stoichiometric compositions of the form  $A_nB$ , where  $n = p^3 - 1$  and  $p$  is an integer. Although there is some experimental evidence for the formation of compounds with the first composition in this series,  $A_7B$ , it seems unlikely that such ordering will occur in many alloy systems. The ground state diagrams of Richards and Cahn (1971) shown in Figs. 6.10 and 6.11 include one cubic superlattice of the f.c.c. structure ( $L1_2$  in Strukturbericht notation) and three of the b.c.c. structure (B2, B32 and DO<sub>3</sub>). Figure 20.9 shows the b.c.c. superlattices in terms of the occupancy of sites on four interpenetrating f.c.c. lattices, from which it is immediately obvious that a mode derived from the usual b.c.c. mode has  $q = 4$  for B2 and  $q = 8$  for B32 and DO<sub>3</sub>. The true, no-shuffle mode of lowest shear for the B2 structure actually has the same  $K_1$  plane as the normal disordered mode, but the direction of  $\eta_1$  is reversed and the shear magnitude is twice as large. A similar "reversed f.c.c. shear" will produce a true twin in the  $L1_2$  superlattice, but a very large shear, four times that of the disordered structure and applied in the same  $\eta_1$  direction, is required to satisfy the geometrical conditions for true twins in either B32 or DO<sub>3</sub>. The B2 and  $L1_2$  twinning modes were discussed by Arunchalam and Sargent (1971), who pointed out that they may be deduced directly from the table of modes published by Bevis and Crocker (1969).

When the superlattice is non-cubic, the new modes derived from different variants of the disordered mode are no longer all equivalent. Some of these derived modes, applied without interchange shuffles, will again change the crystal structure and so may be described as pseudomodes. Other variants, however, will produce true twins in the ideal superlattice. In a detailed discussion based on the Bevis-Crocker theory, Christian and Laughlin (1987, 1988) have enumerated the modes for all non-equivalent variants of the non-cubic structures predicted by Richards and Cahn. Two types of true twin are discussed, namely type I/II, where the direct mode gives a type I orientation and the

conjugate mode a type II, or vice versa, and “combined”, where these two orientations are equivalent. The term “combined” is used rather than “compound” as all the modes are necessarily compound in the original sense that all the twinning elements are rational.

### 86. OPERATIVE TWINNING MODES IN METALLIC STRUCTURES

We may now apply the theory of the previous section to the prediction of the most likely twinning modes in any structure, assuming these modes to be governed by the magnitude of the lattice shear and the complexity and magnitude of the shuffles. Equation (85.13) may be rewritten as

$$s^2 = 4 \frac{|\mathbf{w}|^2}{(\mathbf{w} \cdot \mathbf{m})^2} - 1 \quad (86.1)$$

Now writing  $s < s_{\max}$ , where  $s_{\max}$  is any chosen maximum shear, and using eqn. (85.14),

$$|\mathbf{w}|^2 < q^2 d^2 \{(s_{\max}/4) + 1\} \quad (86.2)$$

which is a condition on the interplanar spacing  $d$  of a set of possible  $K_1$  planes and the shortest lattice vector  $\mathbf{w}$  between two such planes  $qd$  apart. The inequality (86.2) reduces to (85.2) for  $q=2$ , and to (85.3) for  $q=4$ , if  $|\mathbf{w}|$  is replaced by  $|\mathbf{b}|$ . This may sometimes be convenient but, as noted in the previous section, the inequality then becomes a necessary condition only (i.e. it may not be a sufficient condition) for the minimum shear on the planes of spacing  $d$  to be less than  $s_{\max}$ . There is no need to investigate separately the conditions for type II twinning, as the possible type I modes with  $K_1$  rational will automatically give the type II modes with  $K_2$  rational.

The inequality (86.2) is independent of any coordinate system but, in using it, care must be taken if a superlattice cell of higher symmetry is used instead of a primitive unit cell. Thus if  $\mathbf{G}, \mathbf{G}^*$  are the metrics of the primitive bases  $\mathbf{A}, \mathbf{A}^*$ , and we represent the  $K_1$  planes by the vector  $\mathbf{k} = k_i \mathbf{a}_i^*$ , where  $k_i$  are integers with no common factor, inequality (86.2) becomes, in subscript form (see Table II),

$$(G_{ij} w_i w_j)(G_{ij}^* k_i k_j) < (w_i k_i)^2 \{(s_{\max}^2/4) + 1\}. \quad (86.3)$$

However, if we use a larger unit cell for convenience, and  $k_i$  and  $w_i$  are still given integral values, we have to introduce the cell factors  $I$  (p. 38). Using Table III, the inequality now becomes:

$$(I')^2 (G_{ij} w_i w_j)(G_{ij}^* k_i k_j) < I^2 (w_i k_i)^2 \{(s_{\max}^2/4) + 1\}. \quad (86.4)$$

It is frequently more convenient to refer the twinning elements to a centred cell, and thus to use the inequality (86.4).

The prediction of the twinning elements for the single lattice structures was discussed on p. 860 where it was found that, for most metallic structures, the normally observed twinning mode is the mode of lowest shear consistent with the absence of shuffles. In the structures for which this statement is valid, the operative mode and its conjugate are

crystallographically equivalent, and the shear is very much smaller than that of any rival "non-shuffle" mode. This is not true for mercury, which is considered separately below.

In some double lattice structures, it is commonly observed that more than one twinning mode may be active during deformation, but almost all of the many experimental observations of deformation twinning in cubic and tetragonal single lattice structures indicate that only the minimum shear modes are operative. However, there are isolated reports of additional twinning modes, for example in b.c.c. iron–beryllium alloys and in b.c.t. iron–nickel–carbon martensites, and there is also the possibility that transformation twinning in martensite (see Chapter 22) may utilize effective twinning modes different from those produced by plastic deformation. Thus, it is now appropriate to consider briefly the application of the Bevis and Crocker general theory to the prediction of other possible modes in single lattice structures, including non-conventional modes which do not satisfy the classical definition of a twin.

It follows from the description on p. 884 that a unimodular lattice correspondence  $\mathbf{C}$  leading to twinning elements  $K_1$ ,  $K_2$ ,  $\eta_1$ ,  $\eta_2$  and  $s$  will also have an associated conjugate mode obtained by interchanging  $K_1$  and  $K_2$  and  $\eta_1$  and  $\eta_2$ . For cubic lattices, the pair of additional modes obtained by interchanging the indices of  $K_1$  and  $\eta_1$  and of  $K_2$  and  $\eta_2$ , and reversing the sign of either the old or the new  $K_2$  and  $\eta_1$ , also have the same shear magnitude although the amount of shuffling required will be different if the lattice is centred. Bevis and Crocker (1969) used the 60 correspondence matrices mentioned on p. 884 as input data to derive the twinning elements of the corresponding modes for cubic lattices and gave examples from all seven classes of correspondence. They published a table showing a selection of their results in the form of 26 different sets of indices derived from the various matrices  $\mathbf{C}$  and  $\mathbf{C}^{-1}$  each giving rise, in the absence of crystallographic degeneracy, to up to four different twinning modes by applying the above permutations of indices. Of the 26 basic modes, 13 are both conventional and compound whilst the other 13 are non-conventional, and 11 of these have four irrational twinning indices of the form  $x \pm y^{1/2}$ , where  $x$  and  $y$  are integers. Many of the  $K_1$  planes in the compound modes are mirror planes so that the shear would reproduce the parent lattice in the same orientation and this operation could thus not be described as twinning. However, in such cases the  $K_2$  plane is generally not a mirror plane so that the conjugate shear represents a possible twinning mode. The plane of shear is always rational for the non-conventional modes, and this is a general feature of cubic lattices; the orientation relation may be described as a rotation about the normal to this plane of shear.

To investigate the possible twinning modes in the six non-cubic crystal systems, Bevis and Crocker first considered the modes which arise from variants of the unit correspondence matrix. The unit matrix itself leads, of course, to zero shear in all systems but nine of its variants obtained by interchanging rows, interchanging columns and changing the signs of rows and columns (see p. 884) satisfy the restriction (85.43) and so may lead to twins in some crystal systems. These nine correspondence matrices are all symmetric and three of them are diagonal; they all satisfy the condition  $\mathbf{C} = \mathbf{C}^{-1}$  and so represent conventional twinning modes with at least two rational elements. The nine independent modes of the triclinic system reduce to four modes in the monoclinic system,

to three modes in the orthorhombic system, to a single mode (excluding shears which restore the original lattice) in the hexagonal system, to single modes in the tetragonal and rhombohedral systems, and of course to no modes in the cubic system.

The procedure was repeated for a slightly more complex correspondence which has 20 variants leading to twinning in triclinic systems, reducing to eight, six, three, two and one independent modes in the monoclinic, orthorhombic, tetragonal, hexagonal and cubic systems respectively. Most of these modes are non-conventional with four irrational elements.

The twinning modes predicted for the f.c.c. and b.c.c. structures by the no-shuffle, minimum shear hypothesis on p. 862 are both derived from the correspondence matrix

$$C = \begin{pmatrix} 0 & 1 & 1 \\ \frac{1}{2} & \bar{1} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \bar{1} \end{pmatrix} \quad (86.5)$$

which leads to the f.c.c. mode, whilst  $C'$  leads to the b.c.c. mode. Shears of magnitude  $2^{-1/2}$  on the  $\{111\}$  plane lead to twins in the simple cubic, f.c.c. and b.c.c. structures but the fractions of parent lattice sites which are sheared directly to twin positions are respectively  $\frac{1}{2}$ , 1 and  $\frac{1}{4}$ , whereas for a twinning shear of the same magnitude on the  $\{112\}$  plane, these fractions are  $\frac{1}{2}$ ,  $\frac{1}{4}$  and 1. Several other examples of f.c.c. or b.c.c. modes which involve no shuffles are found among the cubic modes listed by Bevis and Crocker but these are all conventional type modes, and non-conventional modes with zero shuffles are only possible if shear magnitudes outside the limits specified on p. 883 are allowed.

All reported instances of twinning in f.c.c. structures have the expected  $\{111\}$  habit plane but, in an experimental study of twinning in iron-beryllium alloys and iron-nickel-carbon martensites, Richman and Conrad (1963) and Richman (1964) found evidence for b.c.c. twins with rational  $\{013\}$  and irrational “ $\{089\}$ ” and “ $\{127\}$ ” habits. However, the iron-beryllium alloys were ordered and twinned copiously on the  $\{112\}$  planes whilst the untempered martensites should have been b.c.t. rather than b.c.c. and, in addition, presumably contained a fine structure of  $\{112\}$  transformation twins. These factors, combined with the small size of most of the anomalous twins, makes the experimental determination of the anomalous modes very difficult and, in a later electron microscopic study, it was concluded that the unusual twin modes observed by Richman in martensite were probably  $\{112\}$  deformation twins with macroscopic habits deviated from the  $\{112\}$  twin plane because of their interaction with the transformation twins. In a later study of iron-beryllium alloys, Green and Cohen (1979) were unable to find evidence of the anomalous modes reported by Richman. Table XVI lists all the predicted b.c.c. modes with  $s \leq 2$  and the modes in which one-half of the lattice points shuffle with  $s \leq 1$ . The  $\{013\}$  habit is seen to be a possible no-shuffle mode, albeit with a rather large shear, but the other two of Richman's reported modes are not among those predicted.

TABLE XVI. PREDICTED AND OBSERVED TWINNING MODES IN SINGLE LATTICE STRUCTURE (CHRISTIAN AND MAHAJAN, 1995)

Structure	$K_1$ { }	$K_2$ { }	$\eta_1$ ( )	$\eta_2$ ( )	$s$	P { }	References to some supporting experimental studies
f.c.c	111	$\bar{1}\bar{1}\bar{1}$	$11\bar{2}$	112	$2^{-1/2}$	$\bar{1}\bar{1}0$	Cu, <sup>1,2</sup> Ag and Cu alloys, <sup>3,7</sup> Ni, <sup>9,10</sup> Al alloys <sup>11,12</sup> and Co Fe alloys <sup>13,16</sup>
b.c.c	112	$\bar{1}\bar{1}\bar{2}$	$\bar{1}\bar{1}\bar{1}$	111	$2^{-1/2}$	$\bar{1}\bar{1}0$	Fe and its alloys <sup>17,31</sup> , Nb and its alloys <sup>32,34</sup> , Mo-Re alloys <sup>35,36</sup> , Mo <sup>40</sup>
(no shuffle)	147	$\bar{1}01$	$\bar{3}\bar{1}\bar{1}$	111	$3^{1/2}/2^{1/2}$	$\bar{1}\bar{2}\bar{1}$	Fe-Be alloys <sup>41,42</sup>
	112	110	$11\bar{1}$	001	$2^{1/2}$	$\bar{1}\bar{1}0$	
	013	415	$5\bar{3}\bar{1}$	$\bar{1}\bar{1}\bar{1}$	$7^{1/2}/2^{1/2}$	$\bar{2}\bar{3}\bar{1}$	
b.c.c	112	$3\bar{3}\bar{2}$	$11\bar{1}$	113	$2^{1/2}/4$	$\bar{1}\bar{1}0$	Fe Ni and Fe Ni C martensites <sup>43</sup> , Fe Ni alloy <sup>44</sup>
( $\frac{1}{2}$ atoms shuffle)	5,8,11	$\bar{1}01$	$5\bar{1}\bar{3}$	111	$6^{1/2}/4$	$\bar{1}\bar{2}\bar{1}$	
	145	$34\bar{1}$	$11\bar{1}$	$\bar{1}39$	$7/58^{1/2}$	$3\bar{2}\bar{1}$	
f.c.t.	013	$0\bar{1}\bar{1}$	031	011	1	$\bar{1}00$	
	011	$0\bar{1}\bar{1}$	$0\bar{1}\bar{1}$	011	$\gamma^1 - \gamma$	$\bar{1}00$	
	111	$11\bar{1}$	$11\bar{2}$	112	$(2\gamma^2 - 1)/2^{1/2}\gamma$	$\bar{1}\bar{1}0$	
b.c.t	011	$0\bar{1}\bar{1}$	$0\bar{1}\bar{1}$	011	$\gamma^1 - \gamma$	$\bar{1}\bar{1}0$	
(no shuffle)	112	$\bar{1}\bar{1}\bar{2}$	$\bar{1}\bar{1}\bar{1}$	111	$(2 - \gamma^2)/2^{1/2}\gamma c^*$	110	
	121	a	b	111	$(2 - 5\gamma^2 + 5\gamma^4)^{1/2}2\gamma$	d	
b.c.t	e	$\bar{1}01$	$5\bar{1}3$	f	$(26 - 29\gamma^2 + 9\gamma^4)^{1/2}4\gamma$	g	
( $\frac{1}{2}$ atoms shuffle)	h	$\bar{1}\bar{1}0$	$5\bar{3}\bar{1}$	111	$(4 + 2\gamma^2)^{1/2}4$	$11\bar{2}$	
f.c.rhombohedral	011	100	100	011	$8^{1/2}c/(1+2c)^{1/2}(1+c)^{1/2}$	$0\bar{1}\bar{1}$	
	j	$\bar{1}\bar{1}\bar{1}$	$\bar{1}\bar{2}\bar{1}$	k	$\frac{(2+8c+22c^2)^{1/2}}{2(l-c)^{1/2}(1+2c)^{1/2}}$	m	Hg <sup>45,47</sup>

Notes: The two f.c.t. modes are linked to their equivalent b.c.t modes. All b.c.c no shuffle modes with  $s \leq 2$  and half-shuffle modes with  $s \leq 1$  are included, but not all the derived b.c.t modes are listed.  $\gamma$  is the axial ratio of the appropriate tetragonal lattice ( $\approx 1.035$  for b.c.t. martensite), and  $c$  ( $\approx -1/7$  for Hg) is the cosine of the rhombohedral interaxial angle. The irrational indices a, b and d-m are:

$a = \gamma^2 - 2, 3\gamma^2, 2 - 3\gamma^2$ ;  $b = -1 - \gamma^2, 3\gamma^2 - 1, 3 - 5\gamma^2$ ;  $d = 2\gamma^2 - 1, 1 - \gamma^2, -\gamma^2$ ;  $e = 14 - 9\gamma^2, 8, 26 - 15\gamma^2$ ;  $f = 5 - 3\gamma^2, 1 + \gamma^2, 5 - 3\gamma^2$ ;  $g = -1.5 - 3\gamma^2 - 1$ ;  $h = 6 - \gamma^2, 10 + \gamma^2, 8\gamma^2$ ;  $j = -1 - 5c, -1 - c, 1 - 3c$ ;  $k = -1 - 7c, -2 - 6c, 1 - c$ ;  $m = 1 + c, -2c, 1 + 5c$ .

\* Predicted minimum shear, no-shuffle modes which have not been observed.

### Key to references

- 1) Blewitt, Coltman and Redman (1957)
- 2) Mahajan, Barry and Eyre (1970)
- 3) Suzuki and Barrett (1958)
- 4) Venables (1961)
- 5) Narita and Takamura (1974)
- 6) Mori and Fujita (1977)
- 7) Mori and Fujita (1980)
- 8) Haasen (1958)
- 9) Robertson (1986)
- 10) Haasen and King (1960)
- 11) Gray (1988)
- 12) Pond and Garcia-Garcia (1981)
- 13) Chin, Hosford and Mendorf (1969)
- 14) Mahajan and Chin (1973a)

## Key to references (continued)

- |   |   |
|---|---|
| 15) Mahajan and Chin (1974)               | 32) Boucher and Christian (1972)                    |
| 16) Mahajan and Chin (1973b)              | 33) Wessel, France and Begley (1961)                |
| 17) Priestner and Leslie (1965)           | 34) McHargue (1964)                                 |
| 18) Le, Bernstein and Mahajan (1993)      | 35) Hull (1963c)                                    |
| 19) Paxton (1953)                         | 36) Mahajan (1972a)                                 |
| 20) Leslie, Hornbogen and Dieter (1962)   | 37) Mahajan (1972b)                                 |
| 21) Mahajan (1969)                        | 38) Mahajan (1975)                                  |
| 22) Altshuler and Christian (1966)        | 39) Mahajan (1971)                                  |
| 23) Zukas and Fowler (1961)               | 40) Mahajan and Bartlett (1971)                     |
| 24) Mahajan (1970)                        | 41) Richman and Conrad (1963)                       |
| 25) Rosenfield, Averbach and Cohen (1963) | 42) Richman (1963)                                  |
| 26) Hull (1961)                           | 43) Rowlands, Fearon and Bevis (1968)               |
| 27) Hull (1963a)                          | 44) Rowlands, Fearon and Bevis (1970)               |
| 28) Hull (1963b)                          | 45) Guyoncourt and Crocker (1968)                   |
| 29) Honda (1961)                          | 46) Crocker, Heckscher, Bevis and Guyoncourt (1966) |
| 30) Terasaki (1967)                       | 47) Abell, Crocker and Guyoncourt (1971)            |
| 31) Edmondson (1961)                      |   |

Further experiments on twinning in cubic iron–nickel and tetragonal iron–nickel–carbon martensites were reported by Rowlands *et al.* (1968). These authors found some b.c.c. deformation twins with the  $\{5, 8, 11\}$  habit of Table XVI. Although all four sets of indices are rational, so that the mode is compound, the orientation relationship is of type II, probably because of a minimum shuffle criterion. With a type II orientation relation, only one-half of the atoms have to shuffle whereas five-sixths of the atoms would have to shuffle to restore the structure in a type I relation. As the  $\{\bar{1}01\}$  plane is a mirror plane, the reciprocal mode is not a true twinning mode in b.c.c.; a shuffle of half the atoms results in an unchanged orientation of the original structure. Rowlands *et al.* (1970), in a further investigation, reported that deformation twins of other than  $\{112\}$  type are extremely rare, but they found some fine (possibly transformation) twins with a  $\{145\}$  habit, which is the conjugate mode to the  $\{013\}$  mode reported by Richman. In contrast, Fearon and Bevis (1974), in a later publication from the same laboratory, reported only  $\{112\}$  transformation twins in a cubic iron–nickel alloy.

Rowlands *et al.* (1968) also tentatively identified two tetragonal derivatives of the “ $\{5, 8, 11\}$ ” mode in the carbon-containing martensite, both of which correspond to conventional type II modes. They point out, however, that non-conventional derivatives of this mode are also possible and there seems no reason why these should not occur in suitable circumstances. Although only one-half of the atoms are sheared to the correct positions, this mode has a smaller shear than the usually observed b.c.c. mode (see Table XVI). A possible reason for its occurrence in martensites is that, in the cubic structure, the  $\{5, 8, 11\}$  twin can propagate undeviated across a  $\{112\}$  twin boundary, and hence across the set of fine parallel  $\{112\}$  twins which are produced by the transformation mechanism. Bevis and Vitek (1970) suggested that a possible reason for the observation of some fine  $\{145\}$  twins in martensite is that a determining factor is the interfacial energy of the coherent  $K_1$  habit plane, rather than the magnitude of the shear, as atomistic calculations indicate that the  $\{145\}$  habit has the next lowest interfacial energy after  $\{112\}$ .

The present position seems to be that the anomalous twinning modes have been identified in the b.c.t. structure with no more certainty than in b.c.c. and the overwhelming majority of observations on both deformation and transformation twins in b.c.t. martensites show only the dominant twinning mode to have the same elements as the b.c.c.  $\{112\}$  mode and a shear  $s = 2^{1/2}(a/c) - 2^{-1/2}(c/a)$ . This mode is derived from the correspondence (86.5) but, as already noted on p. 862, it is also the same mode as that deduced for the f.c.t. structure. Any b.c.t. structure may, of course, alternatively be regarded as f.c.t. with  $a_{\text{fct}} = 2^{1/2}a_{\text{bct}}$  and when this change is made the  $\{112\}$  plane of the b.c.t. structure becomes the  $\{101\}$  plane of the f.c.t. structure, with similar changes of the other twinning elements as shown in Table XVI. When face-centred indices are used, this twinning mode arises from four of the variants of the unit correspondence matrix and when the axial ratio is made equal to unity, so that the structure is f.c.c., the shear becomes zero and the twinning mode ceases to exist. When the structure becomes b.c.c., in contrast, the shear is not zero but  $2^{-1/2}$  and the usual b.c.c. crystallography applies. As pointed out on p. 862, there are equivalent relations between the b.c.t.  $\{011\}$  mode and the f.c.c.  $\{111\}$  mode, and these are also shown in Table XVI. In general, it is convenient to choose the unit cell of a centred tetragonal structure so that the axial ratio differs as little as possible from unity. Thus ferrous martensites with interstitial solutes are described as b.c.t. and have twinning shears of approximately  $2^{-1/2}$ , as the  $\{011\}$  b.c.t. low shear mode apparently does not operate, whereas indium and its alloys are described as f.c.t. and have very small twinning shears.

Now consider twinning in solid mercury, which has a rhombohedral structure and is notable, as pointed out on p. 862, as the only other known example of a single lattice structure in which the no-shuffle, minimum shear mode is not the operative twinning mode. In contrast to the other single lattice structures, the predicted lowest shear mode and its conjugate are not crystallographically equivalent for the rhombohedral structure, so that there are two possible  $K_1$  planes, namely (001) and (110) for shears of lowest magnitude. Early observations (Andrade and Hutchings, 1935) suggested that only the (110) plane is operative, but it now appears that these results were incorrect and that the true mode is of type II with an irrational habit plane close to (135) (Crocker *et al.*, 1966; Guyoncourt and Crocker, 1968; Abell *et al.*, 1971). The elements of this mode are given in Table XVI; it involves no shuffles and has the second smallest shear, which is nevertheless appreciably larger than the shear of the (001) (110) conjugate pair (0.63 and 0.46 respectively for mercury). The conjugate type I mode is seen from Table XVI to have a (111)  $K_1$  plane, but apparently does not occur. Crocker speculates that this may be because the slip plane is also (111) and points out that the same reason was previously advanced for the non-appearance of the (001) mode at a time when both the slip plane and the twinning plane had been incorrectly determined. It is also interesting to note that the observed mode has the same correspondence (86.5) as the f.c.c. and b.c.c. modes, whilst the unobserved minimum shear modes are derived from variants of the unit correspondence matrix.

The twinning modes of the pure components which have single lattice structures are normally also found for essentially disordered solid solutions based on these single lattice

TABLE XVII. POSSIBLE TRUE TWINNING MODES IN CUBIC SUPERLATTICES AFTER CHRISTIAN AND LAUGHLIN (1988)

Mode No.	$S$	$K_1$	$K_2$	$\eta_1$	$\eta_2$	$s^2$	True twin in	Some supporting experimental results
<i>(a) Modes without shuffles</i>								
1.3	(110)	( $\bar{1}\bar{1}$ )	(001)	$[\bar{1}12]$	$[\bar{1}\bar{1}0]$	2	L1 <sub>2</sub>	Cu <sub>3</sub> Au <sup>1, 2</sup> and Ni <sub>3</sub> (Al, Ti) <sup>3, 4</sup> Ti-Ni and Ti-Fe-Ni alloys <sup>5</sup> Fe <sub>3</sub> Al(DO <sub>3</sub> ) <sup>6</sup>
1.3 <sup>T</sup>	(110)	( $\bar{1}12$ )	(110)	$[\bar{1}\bar{1}1]$	$[001]$	2	B2	
1.9	( $\bar{1}\bar{1}0$ )	( $\bar{1}12$ )	(001)	$[\bar{1}\bar{1}1]$	$[\bar{1}\bar{1}0]$	8	B2, B32, DO <sub>3</sub> , L1 <sub>2</sub>	
1.9 <sup>T</sup>	( $\bar{1}\bar{1}0$ )	(111)	(110)	$[\bar{1}\bar{1}2]$	$[001]$	8	B2, B32, DO <sub>3</sub> , L1 <sub>2</sub>	
<i>(b) Modes with 50% (non-interchange) shuffles</i>								
2.3 <sup>T</sup>	(110)	( $\bar{1}\bar{1}4$ )	( $\bar{1}\bar{1}0$ )	$[2\bar{2}1]$	$[00\bar{1}]$	$\frac{1}{2}$	B2	Ti-Ni and Ti-Fe-Ni alloys <sup>5</sup>
1.2	(001)	( $\bar{1}20$ )	(100)	$[210]$	$[0\bar{1}0]$	1	B2, L1 <sub>2</sub>	
2.5	(001)	( $\bar{1}30$ )	(110)	$[310]$	$[\bar{1}10]$	1	B2, B32, DO <sub>3</sub>	
1.3	(110)	( $\bar{1}\bar{1}1$ )	(001)	$[\bar{1}12]$	$[\bar{1}\bar{1}0]$	2	B2, B32, DO <sub>3</sub>	
1.3 <sup>T</sup>	(110)	(112)	(110)	$[\bar{1}\bar{1}1]$	$[001]$	2	L1 <sub>2</sub>	

**Key to references**

- 1) Mikkola and Cohen (1966)
- 2) Chakraborty and Starke (1975)
- 3) Guimier and Struden (1970)
- 4) Kear and Oblak (1974)
- 5) Goo, Duerig, Melton and Sinclair (1985)
- 6) Guedo and Rieu (1978)

solvents. As emphasized on p. 885, however, if sufficient long-range order develops, so that the structure may be regarded as a perfect or imperfect superlattice of the disordered structure, the ordinary twinning mode may become a pseudomode which, in the absence of interchange shuffles, will produce incorrect ordering in the sheared lattice. For superlattice structures which retain cubic symmetry, all variants of the normal mode become pseudomodes of the superlattice. The true modes are listed in Table XVII, but only the L1<sub>2</sub> mode seems actually to occur. The B2 mode, which is also included among the possible b.c.c. modes of Table XVI, has the same shear as the L1<sub>2</sub> mode but has not been observed. Some B2 alloys are known to form pseudotwins in preference to the higher-shear, true mode, whilst in others an alternative twinning mode of lower shear in which 50% of the atoms undergo non-interchange shuffles has been found experimentally. Paxton (1994) has calculated using density functional theory in the local density approximation that for energetic reasons it is impossible to form a pseudotwin of a fully ordered B2 alloy of equiatomic composition.

The pseudomode in B2 is formally a martensitic transformation from the simple cubic structure with space group  $Pm\bar{3}m$  to an orthorhombic structure of space group  $Cmmm$ . The product structure is sometimes erroneously stated to be tetragonal because two of the axes of the orthorhombic cell are equal in the idealized case of no change of lattice parameters on ordering. Relaxation of the parameters will lead, in principle, to three unequal

axes but this may be difficult to detect. Because the pseudomode and the true mode of lowest shear have the same  $K_1$  plane, it is thus essential to obtain experimental evidence of either the symmetry of the product or the magnitude of the shape deformation in order to establish which mode is operating. For a two-phase  $\alpha + B2$  structure in alloys of approximate composition  $Fe_3Be$ , Green and Cohen (1979) showed that there is indeed the anticipated change of symmetry in the "twins" formed in the B2 regions, and they linked this observation with the pseudoelastic behaviour of these alloys (see Section 104). Of course, their alloys of  $Fe_3Be$  composition cannot have had fully ordered B2 structures.

Iron–beryllium alloys with compositions near  $Fe_3Be$  were formerly reported to form a  $DO_3$  superlattice on ordering. Several authors have stated that deformation twinning is impossible in this structure, but Fig. 20.9 and Table XVII show that *geometric* twinning without shuffles may readily be defined. However, the very large shear makes this twinning mode very improbable in practice, and there are no reported observations of deformation twinning in either the  $DO_3$  or the B32 structures. Rather similar but less complete results have been found for iron–aluminium alloys variously reported to have either the B2 or the  $DO_3$  structure. Cahn and Coll (1961) found that alloys with less than 50% long-range order form pseudotwins, but that twinning was suppressed in more highly ordered alloys. Guedo and Rieu (1978) obtained evidence for twinning and detwinning in alloys with the B2 structure and superelastic effects in alloys with the  $DO_3$  structure, but it is uncertain whether or not this latter effect is due to pseudotwinning.

As an alternative to a pseudomode or a high shear mode, it is possible that a twin might form in a superlattice by a mode which requires non-interchange shuffles. The first evidence for such a mode was given by Goo *et al.* (1985) for titanium–nickel and titanium–iron–nickel alloys with the B2 structure. The mode, which involves shuffles of half the atoms ( $q=4$  in both the b.c.c. and simple cubic lattices), has the same shear magnitude as the disordered b.c.c. mode but  $K_1 = \{114\}$  and  $\eta_1 = (2\bar{2}1)$ ; the full indices are shown in Table XVII. This table also gives the other possible modes with 50% shuffles, all of which have higher shear magnitudes. The shuffles involved are relatively simple and have been discussed by Goo *et al.* and by Christian and Laughlin. Assuming that all atoms are displaced by the shear (i.e. the motif units are *not* treated as rigid), the atoms which must shuffle are alternately A and B on successive planes of shear and are contained in alternate  $K_2$  planes normal to the plane of shear. Goo *et al.* suggested that these atoms are all displaced in the same direction [Fig. 20.10(c-i)]; alternative possibilities in which the atoms move in opposite directions in successive  $K_2$  shuffle planes or in successive rows of one shuffle plane are shown in Fig. 20.10(c-ii) and (c-iii).

In contrast to B2, pseudotwinning has not been reported in  $L1_2$  structures, but there is evidence for the true mode of Table XVII in  $Cu_3Au$  alloys (Mikkola and Cohen, 1966; Chakraborty and Starke, 1975) and in microtwins in  $Ni_3(Al,Ti)$   $\gamma'$  phases (Guimier and Struden, 1970; Kear and Oblak, 1974). In their work on  $Cu_3Au$ , Chakraborty and Starke found that the true mode was observed only in alloys with nearly complete long-range order, whilst disordered or partially ordered alloys formed twins with the usual f.c.c. mode. The f.c.c. type twinning was also observed in highly ordered alloys tested in compression to relatively high strains, which would have reduced the initial long-range

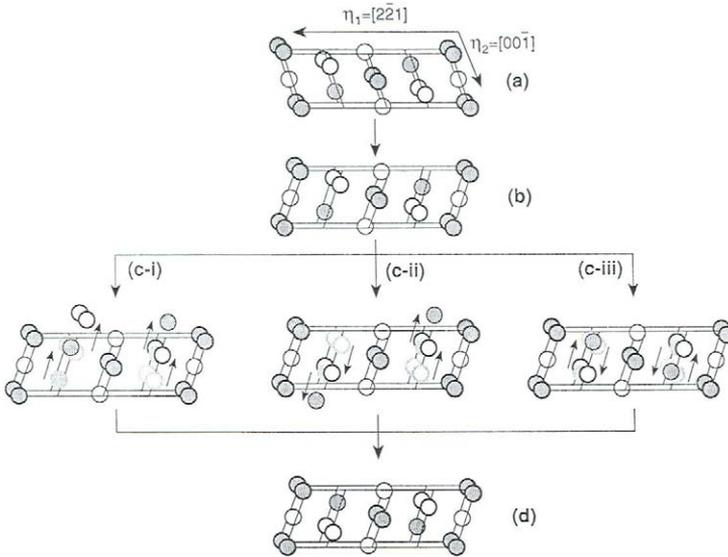


FIG. 20.10. Schematic illustration of the  $\{114\}$  shuffle mode in the B2 structure (after Christian and Laughlin, 1988). *B* and *A* atoms are shown by shaded and open circles respectively. (a) Parent structure; (b) sheared structure; (c-i) (c-ii) (c-iii) alternative shuffles; (d) combined orientation twin.

order. In this case, the twins formed only at applied shear stresses an order of magnitude larger than those required to produce true twins of the ordered structure.

For non-cubic superlattices of the f.c.c. structure, the variants of the usual f.c.c. mode may be divided into six pairs, each comprising a mode and its conjugate and having one of the six cubic  $\{110\}$  planes as the plane of shear. Four of the superlattices listed by Richards and Cahn (1971) were tetragonal, with either two or, in the case of  $\text{Ni}_4\text{Mo}$ , four f.c.c. mode pairs which give true twins in the superlattice. There are five, three, three and three possible true mode pairs for the orthorhombic superlattice ( $\text{Pt}_2\text{Mo}$ ), the two (predicted but unobserved) monoclinic superlattices and the rhombohedral ( $L1_1$ ) superlattice respectively. Experimental results have been reported for the  $L1_0$ ,  $\text{Pt}_2\text{Mo}$ ,  $\text{DO}_{22}$  and  $D1a$  structures.

The most complete investigation of twinning in the tetragonal  $L1_0$  structure is that of Shechtman *et al.* (1974) for a Ti-Al alloy. The specific variants of the  $\{111\}$   $\{11\bar{2}\}$  cubic twin which formed during deformation were identified and shown to correspond to the true modes of the superlattice, and the shear magnitude was also measured. There are several experimental investigations of the  $\text{CuAuI}$  superlattice with the  $L1_0$  structure, among which we may mention that of Pashley *et al.* (1969), who concluded that  $\{111\}$  twinning is an important deformation mechanism and speculated that the structure may be changed in some of the twins.

Hansson and Barnes (1965) and Pashley *et al.* pointed out that the structure produced by pseudotwinning of  $L1_0$  has a single set of  $\{111\}$  cubic planes which are alternately occupied by atoms of each species, so that this structure is effectively that of the

$L1_1$  (CuPt) superlattice. The reverse is also true: the pseudomode of  $L1_1$  yields the  $L1_0$  structure. Unfortunately, there are no experimental results on twinning in  $L1_1$ .

Except at high temperatures, twinning is frequently the major deformation mechanism in alloys with the tetragonal  $DO_{22}$  structure, and it has been studied extensively in  $Ni_3V$  (Vandershaeve and Sarrazin, 1977; Vandershaeve *et al.*, 1979) and in  $Al_3Ti$  (Yamaguchi *et al.*, 1987ab.) The deformation always utilizes only the four true modes (two mode pairs) derived from disordered f.c.c. modes.

Finally, twinning has been established, but not fully investigated in the  $D1a$  structure of  $Ni_4Mo$  (Nesbitt and Laughlin, 1980) and in certain nickel–molybdenum–chromium alloys with the orthorhombic  $Pt_2Mo$  structure (Tawancy, 1981). The true twins in  $Ni_4Mo$  are type I–type II in each conjugate pair whereas, in the  $Ni_3Pt$  structure, one mode pair gives true twins with combined orientations and four pairs give type I–type II orientations. In both structures, it seems probable that true twins are formed but the detailed crystallography was not established.

Now consider the double lattice structures, the most important of which is the h.c.p. structure with axial ratios differing by varying amounts from the ideal  $(8/3)^{1/2}$ . The metals cadmium and zinc, with the rather high axial ratios of 1.886 and 1.856, are usually considered to form a separate subgroup, the remaining metals having axial ratios ranging from slightly smaller than that for ideal close packing (cobalt and magnesium) down to 1.568 (beryllium). All the metals twin on the  $\{10\bar{1}2\}$  planes, but most of the low axial ratio metals have also been reported to twin on several other planes. Additional  $K_1$  planes reported for titanium, for example, are  $\{11\bar{2}1\}$ ,  $\{11\bar{2}2\}$ ,  $\{11\bar{2}3\}$  and  $\{11\bar{2}4\}$ , and additional modes for magnesium include  $\{10\bar{1}1\}$ ,  $\{30\bar{3}4\}$ ,  $\{10\bar{1}3\}$  and  $\{10\bar{1}4\}$ . Some of these “anomalous” twinning modes are well established; in others, the identification of habit plane traces as twins is open to some doubt. Note that the experimental determination of all the elements of a given mode is not possible unless the shear is measured, and this is rather difficult.

Early work showed that the full description of the  $\{10\bar{1}2\}$  mode is

$$\begin{aligned} K_1 \equiv K_2 &= \{10\bar{1}2\}; & \boldsymbol{\eta}_1 \equiv \boldsymbol{\eta}_2 &= \langle 10\bar{1}1 \rangle; \\ \text{Plane of shear} &= \{1\bar{2}10\}; & s &= (\gamma^2 - 3)/3^{1/2}\gamma \end{aligned} \quad (86.6)$$

where  $\gamma$  is the axial ratio. This is one of the two hexagonal modes derived from variants of the unit correspondence matrix applied to the orthohexagonal basis. Crocker used eqn. (86.3) to make a systematic investigation of the low-shear, simple-shuffle modes and, for ideal  $\gamma$ , he listed 11 such modes with  $s < 1$  and  $q \leq 4$ . In a later table published by Crocker and Bevis (1970), 15 possible modes satisfying these restrictions are given for the axial ratio of titanium. For all likely values of  $\gamma$ , the  $\{10\bar{1}2\}$  mode gives much the lowest shear, and there can be little doubt that this is an important factor in the universal observation of this mode. There is also presumably a lower limit to the shear magnitude which can be effectively utilized in twinning; for  $\{10\bar{1}2\}$  twins, for example,  $s$  becomes zero when  $\gamma = 3^{1/2}$  and the shear direction reverses as  $\gamma$  passes through this value, as in magnesium–cadmium alloys. Experiments show that no  $\{10\bar{1}2\}$  twins form near the critical composition (Stoloff and Gensamer, 1963).

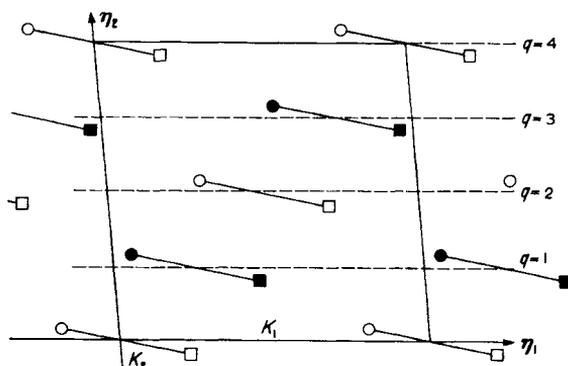


FIG. 20.11. Projection of h.c.p. structure onto plane of shear  $\{1\bar{2}10\}$  of a  $\{11\bar{2}0\}$  twin. Open and filled circles represent atoms of one set of lattice sites on two successive  $\{1\bar{2}10\}$  planes. Squares similarly represent atoms from the second set of lattice sites on the same two  $\{1\bar{2}10\}$  planes.

A projection of the h.c.p. structure onto the plane of shear is shown in Fig. 20.11, from which we see that this is a  $q=4$  mode. It is obvious by inspection that there is only one reasonable choice for the motif unit. The structure is formed by the stacking of two planes of type  $\{1\bar{2}10\}$  in different relative positions, half of the atoms of each lattice being contained in each plane. The motif unit thus lies in the plane of shear, which is a mirror plane of the h.c.p. structure, so that all four orientation relations of eqn. (85.4) are equivalent, and there is no distinction between a type I and a type II twin. As in the f.c.c. and b.c.c. modes, this mode is crystallographically equivalent to its own conjugate.

Comparison of Fig. 20.11 with Fig. 20.8 shows that the probable shuffle mechanism is a degenerate case of Fig. 20.8(e) or 20.8(f), these two mechanisms being identical when the motif unit lies in the plane of shear. The additional atom movements in the shuffles are thus all parallel to  $\eta_1$  or normal to  $K_1$  if the motif unit is regarded as rigid, or parallel to  $\eta_1$  and  $\eta_2$  if the shear is applied homogeneously to all the atom sites. The twinning mode is favoured not only by the low shear, but also by the very simple shuffle mechanism.

The predicted twinning mode with the next lowest shear for  $q \leq 4$  has a  $K_1$  plane of type  $\{22\bar{4}1\}$ , and has not been observed in any h.c.p. metal. The third smallest shear is a  $q=4$ , type I twin with  $K_1 = \{10\bar{1}1\}$ ,  $\eta_2 = \{41\bar{5}3\}$  and  $K_2$  and  $\eta_1$  irrational. This mode is important in the theory of martensite crystallography as it specifies the relationships between the two product lattices in some h.c.p. phases produced by transformation from high temperature b.c.c. phases. Such a transformation occurs, for example, in many titanium- and zirconium-based alloys and the h.c.p. plates are then often finely twinned on a single set of  $\{10\bar{1}1\}$  planes. As discussed in detail in Chapter 22, the reasonable assumption that the two product orientations have crystallographically equivalent correspondences with the parent lattice then fixes the  $\eta_1$  direction and hence the equivalent twin mode (or its conjugate; both have been observed). It is then highly probable that one mechanism for deformation of the h.c.p. product involves the displacement of the  $\{10.1\}$  transformation twin boundaries, as in other martensitic structures.

Deformation twinning on  $\{10\bar{1}1\}$  has also been observed in magnesium but single twins of this type are rare and there may be confusion between genuine  $\{10\bar{1}1\}$   $K_1$  planes and habits which have been described as  $\{30\bar{3}4\}$  but which are probably associated with a double twinning process. Reed-Hill (1960) suggested a  $\{10\bar{1}1\}$  low shear deformation mode, quite different from the  $q=4$  transformation twinning mode. This new mode is compound, and is the reciprocal of a  $\{10\bar{1}3\}$  twinning mode previously noted by Reed-Hill and Robertson (1957); it has  $K_1 = \{10\bar{1}1\}$ ,  $K_2 = \{10\bar{1}\bar{3}\}$ ,  $\eta_1 = \langle 10\bar{1}\bar{2} \rangle$ ,  $\eta_2 = \langle 30\bar{3}2 \rangle$ ,  $s = (4\gamma^2 - 9)/4(3)^{1/2}\gamma$  and  $q=8$ . The magnitude of the shear is 0.138 for magnesium, which is appreciably smaller than that for the  $q=4$  mode. Although only one-quarter of the lattice points are carried to their correct twin positions by the shear, the atomic shuffles are simplified as the plane of shear  $\{1\bar{2}10\}$  is a mirror plane which contains the motif unit, as in the  $\{10\bar{1}2\}$  twins already discussed. An alternative compound mode was suggested by Hall (1954), using the plane of shear projection method, but the shear ( $s \sim 1.07$ ) is improbably large.

The observed  $\{10\bar{1}3\}$  twins in magnesium could in principle represent a type I twin with two irrational elements and  $q=4$  but, although the shear for this mode is reasonably small, a more probable mode, as just discussed, is the reciprocal of the  $\{10\bar{1}1\}$ ,  $q=8$  mode. (A  $q=11$  mode, which has also been suggested, has improbably complex shears.) Experimental observations on magnesium indicate, however, that both  $\{10\bar{1}3\}$  and  $\{10\bar{1}1\}$  twins are often components of double twins rather than single twins. Reed-Hill (1960) found that a twin plate or band appeared to form by double twinning on  $\{10\bar{1}3\}$  and  $\{10\bar{1}2\}$  over most of its length, except at the tips where there was single  $\{10\bar{1}3\}$  twinning. The tips are thus separated from the rest of the twin by internal  $\{10\bar{1}2\}$  interfaces, which presumably move outwards as the twin grows. Similar observations were made for (smaller) double twins with habits near  $\{10\bar{1}1\}$ , in agreement with a suggestion by Couling *et al.* (1959) that observed  $\{30\bar{3}4\}$  twins in magnesium are actually double twins of  $\{10\bar{1}1\} + \{10\bar{1}2\}$  type.

The predicted twinning elements of the equivalent simple shear mode for the  $\{10\bar{1}3\} + \{10\bar{1}2\}$  double twinning are all irrational, but experimental results fit theoretical predictions (Crocker, 1962) quite well. The measured angle between the basal plane and the habit plane, for example, is  $29^\circ$  compared with predicted values of  $32^\circ$  for the  $\{10\bar{1}3\}$  twin alone and  $26-34^\circ$  for the double twin, on the assumption that the latter has a simple shear relation to the matrix produced by simultaneous operation of the two twinning shears. (The alternative possibility [see p. 880] of retwinning a previously formed  $\{10\bar{1}3\}$  twin would require plastic accommodation in the matrix which would slightly rotate the habit to  $32-17'$  from the basal plane.) The small discrepancy of  $2\frac{1}{2}^\circ$  was attributed by Crocker to accommodation effects and he also showed that the expected orientation relations and shear (0.258) of the double twinned region are in good agreement with experiment. However, this is not true for the suggested  $\{30\bar{3}4\}$  double twins.

Two sets of indices were originally suggested tentatively for single  $\{30\bar{3}4\}$  twins; those given by Reed-Hill and Robertson (1957) have  $q=10$ , and those by Couling and Roberts (1956) have  $q=4$ , but also have a very large shear. As there is no mode with  $q \leq 6$  and

$s < 1$ , it seems highly probable that these bands are actually double twins and the combination  $\{10\bar{1}1\} + \{10\bar{1}2\}$  is clearly indicated. However, although the experimental orientations are reasonably consistent with the theory, the measured habit planes deviate by about  $15^\circ$  from those predicted. (This discrepancy is reduced to  $\sim 6^\circ$  if a  $\{10\bar{1}1\}$  twin is assumed to form first and then to retwin.) The results may indicate that the  $\{30\bar{3}4\}$  twins form by a more complex double twinning mechanism, or may simply reflect the experimental difficulties with such small twins; no further work to resolve the discrepancy appears to have been published.

The remaining observed h.c.p. modes are  $\{11\bar{2}1\}$  and  $\{10\bar{1}4\}$ , which both have  $q \leq 4$  and  $s \leq 1$ , and  $\{11\bar{2}2\}$ ,  $\{11\bar{2}4\}$  and  $\{11\bar{2}3\}$ , which are all unpredicted within these ranges of  $q$  and  $s$ . The  $\{10\bar{1}4\}$  observations are rather doubtful, but the other modes, and especially  $\{11\bar{2}1\}$ , are important in the deformation of several h.c.p. metals. We note that  $\{11\bar{2}1\}$  is the only mode listed in Table XVIII with  $q=2$ , i.e. it is the only h.c.p. mode in which all the lattice points are carried to their correct positions by the shear so that *lattice* shuffles are not required. Experimental values for the shear in rhenium (Smith, 1960) and zirconium (Sokurskii and Protsenko, 1958; Reed-Hill *et al.*, 1963; Westlake, 1963) generally agree with the predicted value of  $\sim 0.6$ , but some confusion was caused by an earlier measurement of  $\sim 0.2$  (Rapperport, 1959), which implied a very high value for  $q$ . It now seems probable that undetected accommodation effects were responsible for the low value of  $s$  obtained by Rapperport.

The plane of shear for the  $\{11\bar{2}1\}$  mode is a mirror glide plane  $\{1\bar{1}00\}$ , so that the two orientation relations I and II are in principle different but may be related by a simple translation of  $\frac{1}{2}c$ . The shuffles required to produce a type I twin all involve equal and opposite displacements of the atoms in a motif pair through a small distance ( $\sim 0.2a$ ) perpendicular to the  $K_1$  plane (assuming the motif unit to be translated rigidly by the shear) as in the  $Y$  mechanism of Fig. 20.2. Crocker and Bevis (1970) suggested that the simplicity and small magnitudes of the shuffles account for the dominance of this mode (together with the smaller shear  $\{10\bar{1}2\}$  mode) in the observed deformation behaviour of titanium.

An interesting feature of the  $\{11\bar{2}1\}$  mode is that it provides one of the few examples in which  $K_2$  and  $\eta_2$  correspond to the observed slip plane and slip direction respectively. It follows from the theory of dislocation interfaces (Chapter 8) that a twin boundary may be regarded formally as a high angle tilt boundary formed by a dense array of lattice dislocations of edge type. The boundary could thus, in principle, be created by the accumulation of a large number of slip dislocations in a local region which then rearrange to form a twin boundary with a consequent lowering of energy. Compared with the low angle case (polygonization), it is required that the dislocations must have a lattice Burgers vector in the  $\eta_2$  direction and their glide plane must be  $K_2$ . According to this model, the twinning mode would be determined by the known deformation properties in glide. A knowledge of these elements does not uniquely define the mode, but an additional assumption about the fraction of lattice sites carried directly into their correct twin positions, or equivalently about the spacing of the dislocations in the tilt boundary, is sufficient to fix all the variables.

TABLE XVIII. PREDICTED AND OBSERVED TWINNING MODES IN h.c.p. STRUCTURES

$K_1$ { }	$K_2$ { }	$\eta_1$ ( )	$\eta_2$ ( )	$s$	$q$	References to some supporting experimental studies
$10\bar{1}2$	$\bar{1}012$	$10\bar{1}\bar{1}$	$\bar{1}01\bar{1}$	$(\gamma^2 - 3)/3^{1/2}\gamma$	4	Mg, <sup>1-4</sup> Ti, <sup>4</sup> Co, <sup>5</sup> Zr, <sup>6</sup> Zn, <sup>4,7</sup> and Be <sup>4</sup>
$2\bar{2}\bar{4}1$	0001	1, 1, $\bar{2}$ , $\bar{1}\bar{2}$	$11\bar{2}0$	$1/2\gamma$	4	
$10\bar{1}\bar{1}$	$10\bar{1}\bar{3}$	$10\bar{1}\bar{2}$	$30\bar{3}2$	$(4\gamma^2 - 9)/4(3)^{1/2}\gamma$	8	Mg <sup>1-3</sup> and Ti <sup>6,8</sup>
$10\bar{1}\bar{1}$	i	i	$4\bar{1}\bar{5}3$	$(4\gamma^4 - 17\gamma^2 + 21)^{1/2}/2(3)^{1/2}\gamma$	4	
$20\bar{2}\bar{1}$	0001	$10\bar{1}\bar{4}$	$10\bar{1}0$	$3^{1/2}/2\gamma$	4	
$11\bar{2}1$	0001	$11\bar{2}\bar{6}$	$11\bar{2}0$	$\gamma^{-1}$	2	Re <sup>1</sup> , Ti <sup>4,10</sup> , Zr <sup>11-14</sup> , Co <sup>5</sup> and graphite <sup>15</sup>
$10\bar{1}\bar{3}$	$10\bar{1}\bar{1}$	$30\bar{3}2$	$10\bar{1}2$	$(4\gamma^2 - 9)/4(3)^{1/2}\gamma$	8	Mg <sup>1</sup>
$10\bar{1}\bar{3}$	i	i	$21\bar{1}\bar{3}$	$(4\gamma^4 - 17\gamma^2 + 21)^{1/2}/2(3)^{1/2}\gamma$	4	
i	i	i	i	$(4\gamma^4 - 21\gamma^2 + 36)^{1/2}/4(3)^{1/2}\gamma$		

{ $10\bar{1}3$ } double twinning)

$13\bar{4}0$	$\bar{1}100$	$7\bar{5}\bar{2}0$	$11\bar{2}0$	$3^{-1/2}$	4	
$13\bar{4}1$	$\bar{1}101$	i	$11\bar{2}0$	$(4\gamma^2 + 3)^{1/2}/2(3)^{1/2}\gamma$	4	
$13\bar{4}2$	$\bar{1}102$	i	$11\bar{2}0$	$(4\gamma^2 + 3)^{1/2}/2(3)^{1/2}\gamma$	4	
$2\bar{2}\bar{4}3$	0001	$11\bar{2}\bar{4}$	$11\bar{2}0$	$3/2\gamma$	4	
$10\bar{1}\bar{4}$	$10\bar{1}0$	$20\bar{2}\bar{1}$	0001	$\gamma/3^{1/2}$	4	
$11\bar{2}2$	$11\bar{2}\bar{4}$	$11\bar{2}\bar{3}$	$2\bar{2}\bar{4}3$	$2(\gamma^2 - 2)/3\gamma$	6	Ti <sup>4,6,8</sup> and Zr <sup>6</sup>
$11\bar{2}4$	$11\bar{2}2$	$2\bar{2}\bar{4}3$	$11\bar{2}3$	$2(\gamma^2 - 2)/3\gamma$	6	
$30\bar{3}4$						
$11\bar{2}3$						

Notes: i denotes an irrational plane or direction. As explained in the text, some of the reported modes have not been confirmed and appear doubtful. For greater clarity, each mode and its conjugate are listed separately. The table includes the eleven predicted modes which have  $s \leq 1$  and  $q \leq 4$  for the ideal axial ratio, and also the more probable  $q = 8$  modes for {1011} and {1013} and the  $q = 6$  modes for {1122} and {1124}.

**Key to references**

- 1) Reed-Hill (1960)
- 2) Couling, Pashak and Sturkey (1959)
- 3) Couling and Roberts (1956)
- 4) Partridge (1967)
- 5) Vaidya and Mahajan (1980)
- 6) Yoo (1981)
- 7) Lavrentyev, Salita and Kazarov (1968)
- 8) Paton and Backofen (1970)
- 9) Smith and Gaunt (1961)
- 10) Odinkova (1967)
- 11) Sokurskii and Protsenko (1958)
- 12) Reed-Hill, Slippery and Buteau (1963)
- 13) Westlake (1961)
- 14) Reed-Hill and Dahlberg (1966)
- 15) Freise and Kelly (1961)

TABLE XIX. PREDICTED AND OBSERVED TWINNING MODES IN DOUBLE LATTICE STRUCTURES (AFTER CHRISTIAN AND MAHAJAN 1995)

Structure	$K_1$ { }	$K_2$ { }	$\eta_1$ ( )	$\eta_2$ ( )	$s$	q	References to supporting experimental studies
Rhombohedral	110	001	001	110	0.27—0.125	1	As, Bi, Sb <sup>1</sup>
b.c. tetragonal	101	$\bar{3}01$	$\bar{1}01$	103	0.113	4	Sn <sup>2</sup>
Diamond	111	$\bar{1}\bar{1}\bar{1}$	$\bar{1}\bar{1}\bar{2}$	112	$2^{-1/2}$	2	Si <sup>3</sup> and GaAs <sup>4-7</sup>
	111	$\bar{1}\bar{1}\bar{3}$	$\bar{1}\bar{1}\bar{2}$	332	$8^{-1/2}$	4	
$\alpha$ -uranium	'176'	111	512	—	0.214	4	8
	112	'172'	312	—	0.227	4	9
	'172'	112	—	312	0.227	4	9
	130	110	310	$\bar{1}\bar{1}0$	0.299	2	9
	121	' $\bar{1}\bar{4}\bar{1}$ '	' $\bar{3}\bar{2}\bar{1}$ '	311	0.329	6	9

Notes: Irrational planes or directions are shown as approximately rational indices in quotation marks. Numerical values of the shear given to avoid complexity. Only observed modes are listed.

#### Key to references

- 1) Gough (1933)
- 2) Tu and Turnbull (1970)
- 3) Pirouz, Chaim, Dahmen and Westmacott (1990)
- 4) Androussi, Vanderschaeve and Lefebvre (1989)
- 5) Rabier and Boivin (1990)
- 6) Boivin, Rasbier and Garem (1990a)
- 7) Boivin, Rasbier and Garem (1990b)
- 8) Lloyd and Chiswick (1955)
- 9) Cahn (1953a)

Bullough's formal theory is closely related to the theory of martensite crystallography (Chapter 22) and may also be linked to the discussion of surface dislocations on p. 365. Figure 20.12 shows hypothetical stages in the formation of a twin. The combination of the lattice-invariant deformation (see p. 322), which is a simple shear on  $K_2$  in the direction  $\eta_2$ , with the lattice deformation, which is a rotation about the direction in  $K_2$  normal to  $\eta_2$ , produces the shape deformation, which is a simple shear on  $K_1$  in the direction of  $\eta_1$ . The more complex case of martensite differs formally in having a lattice deformation which is not simply a rotation, and the dislocations needed to ensure compatibility of lattice and shape deformations are then not pure edges. In the application to twinning, however, the dislocation content of the tilt boundary is so high that it loses physical significance; if all the lattice points are related by the twinning shear, there is one lattice dislocation on each  $K_2$  lattice plane intersecting the twin interface. This very high dislocation content may alternatively be regarded as zero, because an equally valid description of the final situation is that the shape and lattice deformations are identical and there is no lattice-invariant deformation (see p. 366). This latter description is the one used in this chapter, but Bullough's theory shows the relevance of the alternative description if the twin really does form by accumulating slip dislocations.

As  $K_2$  and  $\eta_2$  hardly ever correspond to slip modes, this approach cannot be generally valid, but it has been considered in relation to one possible mode in the diamond structure

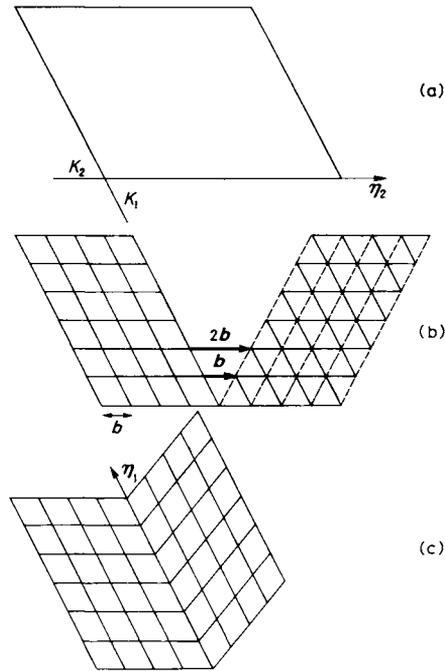


FIG. 20.12. Diagram illustrating the crystallography of Bullough's model of twinning. (a) Original lattice. (b) Part of the original lattice is given a lattice-invariant shear on the  $K_2$  plane; this requires a dislocation of Burgers vector  $\mathbf{b}$  on each lattice  $K_2$  plane as shown. The solid lines outline the original unit cells of the structure; the broken lines show alternative, equivalent cells in the sheared portion. (c) Combination of the lattice-invariant shear with a rotation about the normal to the plane of shear produces the twin. The total shape deformation is a simple shear on the  $K_1$  plane in the  $\eta_1$  direction.

(see below) and also in the interesting case of hexagonal graphite which forms  $\{1\bar{1}2\}$  twins with a much lower theoretical shear (0.367) than titanium because of its high axial ratio ( $\gamma=2.725$ ). Freise and Kelly (1961) found that deformation of graphite frequently produced bend planes (tilt boundaries) of varying angle in addition to genuine twins, thus lending some support to the Bullough model. However, it does not seem possible to assign much more than formal significance to the dislocation model of the completed twin boundary, as the dislocations must lose their individual identities. Freise and Kelly used the Read-Shockley formula for grain boundary energy to deduce that the "dislocations" in the boundary will dissociate into partial dislocations, there being one partial on each atomic  $K_2$  plane instead of one lattice dislocation on each lattice  $K_2$  plane. It is obvious that this statement is only meaningful in a formal sense, where it may be interpreted to show that it is possible to have an interface in a twin of this type in which all the atomic positions match. Statements which are sometimes made that in this type of twinning the shear is on the  $K_2$  plane rather than the  $K_1$  plane focus attention on the dislocation aggregation mode, but are rather confusing because a proper distinction is not being made between the different kinds of deformation shown in Fig. 20.12.

There are no other observed h.c.p. modes for which  $s < 1$  and  $q \leq 4$  but it appears to be well established that  $\{11\bar{2}2\}$  and  $\{11\bar{2}4\}$  twins occur in some metals. In fact, if  $q = 6$ , these two planes are conjugate twinning planes of a single mode, as first pointed out by Kiho (1958), and the shear is only  $\sim 0.22$  for titanium and zirconium, although it would exceed 0.5 if this mode were operative in cadmium or zinc. With  $q = 6$ , the shuffles are necessarily rather complex, but are considerably simplified in each case because there is a motif unit in the composition plane. Type I and type II twins are possible in principle but, as the plane of shear is the mirror glide plane  $\{1\bar{1}00\}$ , the two orientations differ only by a relative translation, which in this case may be regarded as parallel to  $\eta_1$ . This introduces additional symmetry; for example, the atoms in a reflection twin are also related to the parent atoms by a twofold screw rotation about  $\eta_1$ . The atomic structure of the various possible h.c.p. twin interfaces and the corresponding twinning dislocations are discussed in the next section; a possible set of shuffles for the  $\{11\bar{2}2\}$  type I twin is shown in Fig. 20.13.

An alternative mode with a  $\{11\bar{2}2\}$  habit (Hall, 1954) has  $K_2 = \{11\bar{2}\bar{2}\}$ ,  $\eta_1 = \langle 11\bar{2}\bar{3} \rangle$ ,  $\eta_2 = \langle 11\bar{2}3 \rangle$ ,  $s = \gamma - \gamma^{-1}$  and  $q = 4$ ; this is, in fact, the second hexagonal mode obtained from variants of the unit correspondence matrix applied to the orthohexagonal basis and, like the  $\{10\bar{1}2\}$  mode [eqns. (86.6)], it has the  $K_1$  and  $K_2$  planes and the  $\eta_1$  and  $\eta_2$  directions crystallographically equivalent. This mode is listed by Crocker and Bevis (1970) but does not appear in Crocker's original list because  $s$  slightly exceeds unity for  $\gamma = 1.63$ , although it is less than unity for the axial ratios of titanium and zirconium. Although fewer atoms would be required to shuffle than in the  $q = 6$  mode, experimental work (see, e.g., Rapperport and Hartley, 1960) shows that the  $q = 6$  mode is preferred, presumably because of the much smaller shear.

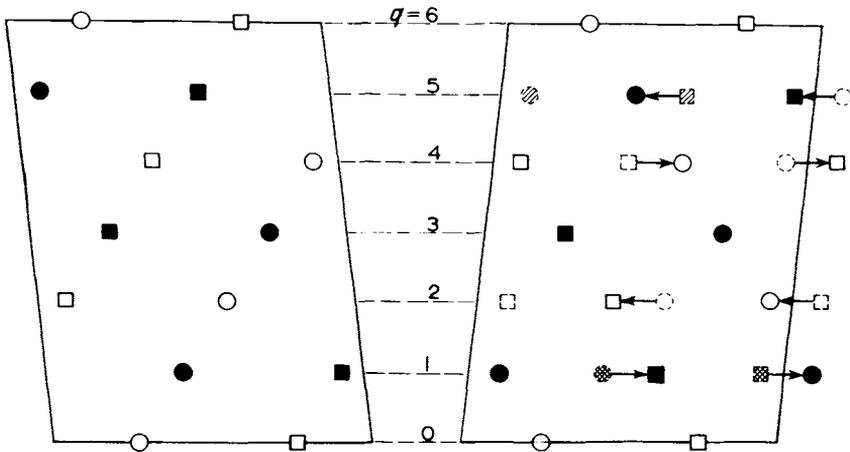


FIG. 20.13. Possible shuffling mechanism for a  $\{11\bar{2}2\}$  type,  $q = 6$ , h.c.p. twin (after Crocker, 1959). The plane of the Figure is  $\{1100\}$  and the symbols are as in Fig. 20.11 except that each  $\{1100\}$  plane contains only one set of lattice sites, so that the four symbols now refer to four successive planes. The dashed and lightly hatched symbols refer to the atom positions after the shear, whilst the full symbols give the atom positions in the twin.

There remains, finally, the observation of  $\{11\bar{2}3\}$  twins. Hall (1954) found a possible mode with  $q=2$ , but this has a very large shear of  $\sim 1.9$ . The lowest shear for this mode with  $q \leq 8$  occurs with  $q=7$ ; another mode suggested by Kiho has  $q=5$  and  $s \sim 0.5$ . Both of these modes would require very complex shuffles. At present there is thus no good description of  $\{30\bar{3}4\}$  and  $\{11\bar{2}3\}$  twinning, and it is possible that the habit plane markings of these types represent rather complex double twinning modes.

Figure 20.14 shows the suggested shuffling mechanisms for four of the operative modes in titanium and Table XVIII gives a selection of observed and predicted twinning modes for the various h.c.p. metals. It does not include the high shear  $\{11\bar{2}\bar{2}\}$  mode mentioned above for which  $q=4$ , and three other unobserved  $q=4$  modes in the Crocker and Bevis list are also omitted. These have  $K_1$  planes of types  $\{2, 8, \bar{1}0, 7\}$  (a compound mode) and  $\{10\bar{1}2\}$  and  $\{10\bar{1}\bar{1}\}$  (type 1 modes).

The metals arsenic, bismuth and antimony are double lattice structures which may be compared to the face-centred rhombohedral structure of mercury as each consists of two interpenetrating face-centred rhombohedral lattices, the origin of the second being displaced along the  $\langle 111 \rangle$  direction with respect to the origin of the first. The observed

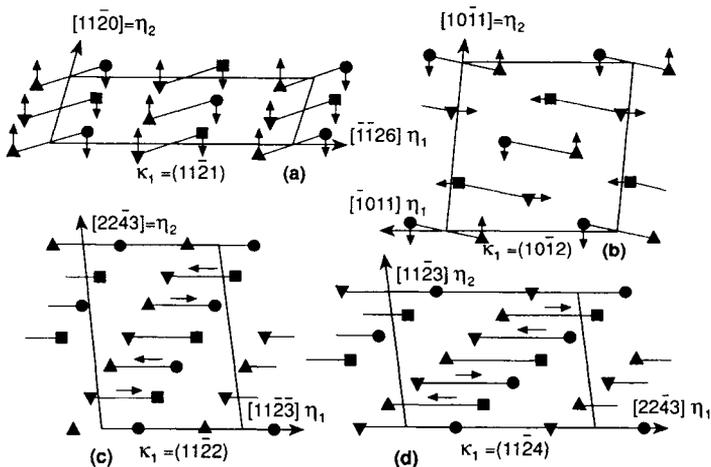


FIG. 20.14. Atomic shuffles associated with (a)  $\{11\bar{2}1\}$ ; (b)  $\{10\bar{1}2\}$ ; (c)  $\{11\bar{2}\bar{2}\}$ ; and (d)  $\{11\bar{2}4\}$  twinning shears in titanium. The  $K_1$  plane and the  $\eta_1$  and  $\eta_2$  directions are labelled in each case and the motif units at each lattice point are indicated. The atoms associated with the two interpenetrating lattices comprising the h.c.p. structure are indicated by circles and squares (first lattice) and by upright and inverted triangles (second lattice). In (a), (c) and (d) the atoms lie in four adjacent  $\{1100\}$  planes, indicated by circles in the plane of the Figure and squares below the plane of the Figure. In (b) the atoms lie in only two adjacent  $\{1\bar{2}10\}$  planes, indicated by circles and upright triangles in the plane of the Figure and by squares and inverted triangles in the next plane at a distance of  $a/2$  below it. The shuffles, assumed formally to precede the shear, are shown by arrows and, in (c) and (d), motif pairs of atoms move together. Shuffle components normal to the plane of the Figure are required in (c) and (d); these are not shown and depend on whether the orientation relationship is of type I or type II.

twinning mode is not, however, equivalent to the operative mercury mode but corresponds instead to the minimum shear mode predicted, but not observed, for mercury. Experimental observations show that the operative  $K_1$  plane is always  $\{110\}$  and the conjugate mode with a  $\{001\}$  habit is not observed. Crocker showed that, with a proper choice of motif unit, the rotation (type II) twin involves much smaller atomic shuffles than the alternatives, and it is presumably preferred for this reason.

The least shear hypothesis also explains the observed twinning modes in the white tin structure, which may be regarded as two interpenetrating b.c.t. lattices with the origin of the second lattice at  $[\frac{1}{2} 0 \frac{1}{4}]$  with respect to the origin of the first. Because of the low axial ratio of the structure ( $\sim 0.55$ ) the mode previously given for martensite (and equivalently for indium) no longer gives the smallest shear. The minimum shear mode is

$$\begin{aligned} K_1 = \{101\}, \quad K_2 = \{\bar{3}01\}, \quad \eta_1 = \langle \bar{1}01 \rangle, \quad \eta_2 = \langle 103 \rangle \\ s = \frac{1}{2}(9\gamma^2 + \gamma^{-2} - 6)^{1/2}, \quad q = 4 \end{aligned} \quad (86.7)$$

Both  $\{101\}$  and the conjugate  $\{\bar{3}01\}$  are observed as  $K_1$  planes, the latter being more frequent. Both possibilities involve quite small and simple shuffles as the motif unit may be chosen to lie in the plane of shear; these represent the degenerate case in which Fig. 20.8(e) is equivalent to Fig. 20.8(f). The preferred plane is that for which the shuffles normal to the  $K_1$  plane are smaller than those parallel to the  $\eta_1$  direction.

The elements carbon (as in diamond), silicon, germanium and grey tin all have the diamond structure in which each atom has four tetrahedrally arranged neighbours around it. The structure is equivalent to two interpenetrating face-centred rhombohedral lattices, the origin of the second being at  $\frac{1}{4} \frac{1}{4} \frac{1}{4}$ . By analogy with the f.c.c. structure, the twin of least shear for  $q=2$  has both  $K_1$  and  $K_2$  planes of type  $\{111\}$ . Jaswon and Dove pointed out, however, that the following  $q=4$  mode is possible in the diamond structure

$$\begin{aligned} K_1 = \{111\}, \quad K_2 = \{11\bar{3}\}, \quad \eta_1 = \langle \bar{1}\bar{1}2 \rangle, \quad \eta_2 = \langle 332 \rangle \\ s = 8^{-1/2}, \quad q = 4 \end{aligned} \quad (86.8)$$

This has quite acceptable shuffles, and one-half of the shear of the  $q=2$  mode.

No observation appears to have been made of the  $\{113\}$  twins which would be conjugate to the above mode, but Churchman et al. (1956) reported a  $\{123\}$  twinning mode. This provided the first application of the theory described on p. 901, and Bullough (1957) showed that the mode could be predicted by using the parent slip system for  $K_2$  and  $\eta_2$ , and not allowing atomic shuffles. Unfortunately, the accompanying shear is improbably large, and later work makes it appear rather doubtful that  $\{123\}$  twins do in fact exist in this structure.

The last twinning modes to be discussed in this section are those of the actinides, which include examples of metallic structures of rather low symmetry. Much work has been published on  $\alpha$ -uranium, which is of considerable interest not only because of its industrial importance but also because it twins on several systems and provided the first known example of a metallic structure in which type II twins form with irrational  $K_1$  planes.

Alpha-uranium has a double lattice structure made up from two interpenetrating base-centred orthorhombic cells, with origins separated by a vector  $[0, -0.31, 0.5]$ . The structure may be regarded as being derived from the h.c.p. structure, and the first theoretical analysis by Frank (1953) showed that one prominent mode with an irrational  $K_1$  plane equal to “{172}” approximately, and its conjugate with  $K_1 = \{112\}$ , which is also an observed mode, may be regarded as the orthorhombic equivalent of h.c.p.  $\{10\bar{1}2\}$  twins. There is another possible orthorhombic mode which also degenerates into  $\{10\bar{1}2\}$ , but this has a larger shear and is not observed. The most frequently occurring twins in  $\alpha$ -uranium have  $K_1 = \{130\}$ , and Frank pointed out that the corresponding h.c.p. plane is a  $\{11\bar{2}0\}$  mirror plane. The above three modes were all discovered in a classic experimental study by Cahn (1953a), who also found another twin with  $K_1 = \{121\}$ . Cahn correctly identified the twinning elements of the  $\{130\}$  (compound),  $\{112\}$  (type I) and “{176}” (type II) modes and also proposed that the  $\{121\}$  habit represents a type I mode with  $\eta_2 = \langle 311 \rangle$ .

Most of the  $\alpha$ -uranium twinning modes were explained by Jaswon and Dove, who also predicted a minimum shear,  $q = 4$ , mode with  $K_1 = \{111\}$  for the type I twin and  $\eta_1 = \langle 512 \rangle$  for the type II twin. This pair of modes may be regarded as the orthorhombic equivalents of the possible h.c.p. type I mode with  $K_1 = \{10\bar{1}1\}$  and  $\eta_2 = \langle 41\bar{5}3 \rangle$  (see p. 897) and the conjugate type II mode. The type II twins with irrational habit plane near “{176}” were discovered in later work by Lloyd and Chiswick (1955) who did not, however, detect the  $\{112\}$  and  $\{121\}$  twins found by Cahn. The Jaswon and Dove theory is unable to distinguish between the observed and unobserved predictions, but a more systematic analysis of the shuffle mechanisms (Crocker, 1965) suggests that the most frequently observed twinning mode, which is compound with  $K_1 = \{130\}$ , is preferred because of its simple shuffle mechanism ( $q = 2$ ). No significance appears to be attached to the fact that the mode is compound; there are actually four possible twinning modes with a lower shear than the  $\{130\}$  mode, including another compound mode, but these all have  $q = 4$ . It is not clear, however, why the reciprocal to the  $\{130\}$  mode, which would have a  $\{110\}$  composition plane, is not observed.

The type I twin of lowest shear has a  $\{111\}$  composition plane, and its reciprocal type II twin an irrational, approximately “{176}”, plane. The shuffles in both cases would be quite large, and only the type II twin has been found. The shuffles for the mode of next lowest shear are simpler, explaining the more frequent occurrence of both the type I,  $\{112\}$ , and the type II, approximately “{172}”, twins. The type II twin has the simpler shuffles, and is observed more frequently than the type I twin. The other two possible modes with  $q = 4$  and a shear less than that of the predominant  $\{130\}$  mode, including the low shear compound mode, are not observed.

The  $\{121\}$  twin found by Cahn, but apparently very rare, cannot be explained by the assumption  $q \leq 4$  used in the above predictions. Crocker points out that, according to the elements assigned to it by Cahn, it is a  $q = 6$  mode, and this does not seem too improbable in view of the  $q = 6$  h.c.p. mode discussed on p. 903. In the uranium case, however, the shuffles would be much more complex as the motif unit cannot be chosen to lie in the  $K_1$  plane.

The twinning modes for the metallic double lattice structures other than h.c.p. are summarized in Table XIX (p. 901).

Mechanical twinning is the major deformation mechanism in many non-metallic materials with complex crystal structures of low symmetry and many atoms to the unit cell. Although the mechanism of such twinning is not well understood, it seems logical to suppose that the operative twinning modes are determined primarily by the shuffle mechanisms, which depend on the structures of the unit cells rather than on lattice geometry. Metallic neptunium provides an interesting example of an intermediate case; it has an orthorhombic unit cell containing eight atoms which, in a (001) projection, can conveniently be represented as binary motif units of two different kinds. Rehtien *et al.* (1971) considered the possible twinning modes to be limited by the condition that the lattice must be restored by the shear, i.e.  $q=1$  or 2. The two lowest shear modes are then

$$K_1 = \{110\}, \quad K_2 = \{1\bar{1}0\}, \quad \eta_1 = \langle 1\bar{1}0 \rangle, \quad \eta_2 = \langle 110 \rangle \quad s = 0.068 \quad (86.9)$$

and

$$K_1 = \{011\}, \quad K_2 = \{01\bar{1}\}, \quad \eta_1 = \langle 01\bar{1} \rangle, \quad \eta_2 = \langle 011 \rangle \quad s = 0.630 \quad (86.10)$$

The very small shear of mode (86.9) suggests that, if the shear magnitude remains an important criterion, this mode will be preferred. Experiments on polycrystalline specimens using colour metallography confirmed that deformation twins form profusely on two habit planes of each grain and an analysis of these observations, combined with a study of habit plane traces in single crystals, showed that all the twins were consistent with the above mode. From a detailed study of the possible shuffle mechanisms, the authors predict that the most likely orientation relationship is a  $180^\circ$  rotation about the  $\eta_1$  direction.

The most complex structure for which a detailed analysis has been attempted is that of  $\alpha$ -plutonium, which has 16 atoms in its monoclinic unit cell. There are many operative slip systems in this metal, but in a few specimens Spriet (1964) also observed deformation twins with two different habits. The habits have not been identified because of the considerable experimental difficulties, but it has been suggested that they are probably (001) and ( $\bar{2}01$ ). In a more complete theoretical analysis Crocker (1971) considered five conjugate pairs of modes with  $q=1$  or 2 and another five pairs with  $q=4$ . The lowest shear mode of the second group has a slightly smaller shear (0.129) than that of the first group (0.159) but its shuffles are much more complex. Crocker examined all the shuffles with the aid of a simpler pseudostructure with only two atoms in its monoclinic unit cell and deduced that only the conjugate pair of modes suggested by Sebilliau, which have minimum shear with no lattice shuffles, is likely to be acceptable. This compound mode is

$$K_1 = (001), \quad K_2 = (\bar{2}01), \quad \eta_1 = [\bar{1}00], \quad \eta_2 = [102] \quad s = 0.159 \quad (86.11)$$

There is no clear distinction between the two conjugate modes but it is suggested that considerations related to the production of twinning dislocations in the interface may favour the ( $\bar{2}01$ ) habit.

From the comparison of theory and experiment made in this section, we can summarize the factors which seem to determine the operative twinning modes in all metallic structures. Bilby and Crocker (1965) suggested that the operative twinning mode should

- (a) have a small shear;
- (b) require only simple shuffles (i.e. have a small value of  $q$ );
- (c) require, if possible, only shuffles of small magnitude; and
- (d) have shuffle displacements parallel to  $\eta_1$  if large shuffles are essential.

Criteria (a) and (b) generally enable the most likely twinning modes to be predicted, but (c) and (d) may have to be invoked in some cases, e.g. to distinguish between a mode and its conjugate.

Laves (1952, 1966) defined a "shuffle parameter" as the ratio of the average shuffle displacement to the interatomic distance, but for any given structure it is not clear that this provides an adequate description of the above considerations. In general, the shuffles might be expected to correlate either with the energy of the interface or the kinetics of its motion, or both, and thus the influence of criteria (b), (c) and (d) on the operative modes is easy to understand in a general way. It is not so clear why a small shear should be preferred as the volume of twins required to produce a given plastic strain field varies as  $s^{-1}$  and thus increases as the twinning shear decreases. However, the stress field of an enclosed lenticular twin of fixed aspect ratio varies as  $s^2$ , so that it should be easier to produce a given strain by a large number of small-shear twins than a smaller number of large-shear twins.

## 87. NUCLEATION AND GROWTH OF MECHANICAL TWINS

In this section, the physical processes by which a deformation twin forms in a solid are discussed. This requires a consideration of the atomic structure of a twin interface and the properties of defects (steps or twinning dislocations) in the interface. Twinning is especially important as a deformation mechanism when there are few operative slip systems, as in the h.c.p. metals of high axial ratio, and it is also very prevalent in the deformation behaviour of b.c.c. metals at low temperatures. The b.c.c. twins are long and thin because of the high shear, and in iron they are known as Neumann lamellae. Metallographic and X-ray investigations in the 1920s (Harnecker and Rassow, 1924; Mathewson and Edmunds, 1928) demonstrated that Neumann lamellae are probably deformation twins, but conclusive proof was not obtained until the work of Kelly (1952) and the twinning elements were first confirmed experimentally by Paxton (1953).

Early evidence of mechanical twinning in f.c.c. metals was discredited, and for many years it was believed that deformation twins do not form in this structure which has identical twinning ( $K_1$ ) planes and slip planes. The work of Blewitt *et al.* (1957) was a major advance as it presented incontrovertible evidence of deformation twinning in copper at 4.2 K and 77 K. Confirmation of f.c.c. twinning came almost immediately from the experiments of Suzuki and Barrett (1958) on silver-gold alloys and those of Haasen (1958)

on nickel, and these pioneering papers were soon followed by evidence that f.c.c. alloys of low stacking fault energies may twin at temperatures up to, and even above, room temperature (Christian *et al.*, 1960; Venables, 1960). Early experimental results are described in the papers from a conference held in 1963 (Reed-Hill *et al.*, 1964) and some of the more recent results were reviewed by Mahajan and Williams (1973) and by Christian and Mahajan (1995).

The understanding of twinning was greatly advanced by a series of experiments on non-metallic materials such as calcite and sodium nitrate carried out in the former Soviet Union by Garber and his co-workers and reviewed by Cahn (1953b). When a stress is applied over a localized region, e.g. with a knife edge, a thin twin wedge may form. If this disappears on removal of the stress, the twin is termed "elastic", but in some circumstances a residual twinned region may remain, and this is called a "stopped elastic twin". The change of shape within the twin wedge is opposed by the constraints of the matrix, and the shear has to be accommodated elastically or plastically within this matrix. However, if the stress is increased sufficiently, a twin may spread across the whole cross-section of a single crystal. It then acquires planar boundaries parallel to the  $K_1$  planes, and is termed a residual twin. Residual twins may be further enlarged by application of a stress; they do not shrink again when the stress is reduced, but may shrink if the stress is reversed.

These observations are not duplicated exactly in metals, where the production of fully elastic twins is rarely observed. The difference appears to be a result only of the presence of other deformation processes acting at low stresses to accommodate the effects produced by the twin in the surrounding matrix. There is, moreover, a close analogy between the phenomenon of elastic twinning in non-metallic materials and that of thermoelastic martensite formation in some alloys (see p. 976). When an elastic twin is formed, there is an equilibrium at all stages between the externally applied stress tending to enlarge the twin, and the internal constraints of the matrix, which oppose the growth of the twin. The matrix stresses then progressively untwin the re-orientated region as the external stresses are reduced. However, if the matrix stresses reach a sufficiently high level, some form of accommodation slip or kinking may occur, and the situation will then usually not be reversible. When the external stresses are removed, the internal stresses will only be partly able to untwin the twinned region, and we obtain a stopped elastic twin. In metals, accommodation may be so complete that there is virtually no untwinning unless the applied stresses are reversed. In the limit, the accommodating deformation may be severe enough to completely destroy coherency between twin and matrix and the twin boundary will then no longer be glissile when a stress of either sign is applied.

A macroscopic twin contained entirely within a matrix crystal almost always has a lenticular shape because of the constraints just mentioned. The ratio of twin thickness to length provides a qualitative guide to the magnitude of the twinning shear, being very small in cases where the twinning shear is large. If an energy minimization principle applies, this ratio can be calculated for elastic twins, and measured values enable the surface energy of the twin interface to be estimated (Christian, 1959). There are many possible ways in which the shape change may be accommodated, depending in

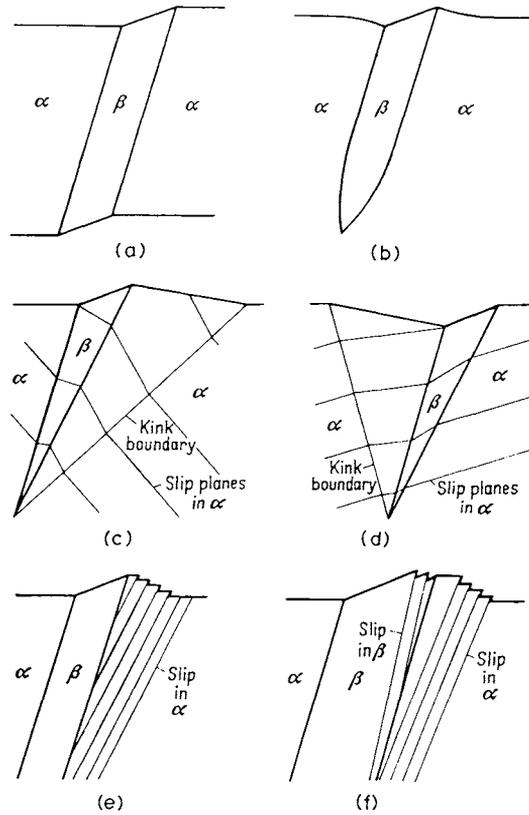


FIG. 20.15. The accommodation of the shape change of twinning: (a) unconstrained twin forming in a single crystal; (b) lenticular twin accommodated elastically in the matrix; (c, d) accommodation by kink boundaries in the matrix; (e) accommodation by slip in the matrix; (f) accommodation by slip in matrix and twin. The Figure applies equally to any displacive transformation.

detail on the crystallography of slip in the structure concerned. Some of the possibilities are illustrated in Fig. 20.15; accommodation by primary kink bands, parallel to the trace of the  $K_1$  plane in the basal plane, is very common in h.c.p. metals and has been extensively studied (Pratt and Pugh, 1952; Moore, 1952; Holden, 1952). In some cases, there are several parallel accommodation kinks of this type, each kink corresponding to an abrupt change in the twin thickness below the surface (Moore, 1955). Secondary accommodation kinks, with traces in the basal plane normal to that of the twinning plane, are also frequently observed; whilst the primary kinks consist of assemblies of basal plane dislocations in materials like zinc, the secondary kinks have been shown to be formed from non-basal plane dislocations (Rosenbaum, 1961).

Figure 20.15 makes it obvious that in some cases the accommodation effects could mask the shape change caused by the twinning shear itself. In principle, there is always an escarpment with the ideal slope of Fig. 20.15(a), but if effects like 20.15(e) or (f) operate on

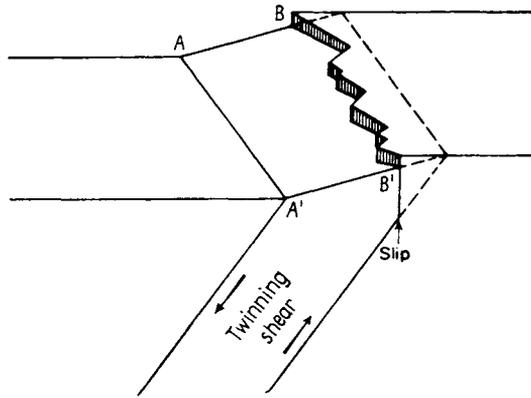


FIG. 20.16. Diagram illustrating the formation of an irregular twin edge (after Hull, 1960). The twin is held up below the surface near  $BB'$ , and the shape deformation causes slip in this region, exposing facets of slip planes.

a sufficiently fine scale, this may not readily be detectable. In the analogous situation in martensitic transformations, there is evidence that the shape change may be obscured by accommodation effects, and this may also be true of some twins.

Twins which entirely cross a single crystal (residual twins in Garber's notation) should have flat parallel boundaries, but examination of the traces in smooth polished surfaces frequently shows that one side is straight whilst the other has a characteristic zigzag marking. This effect was examined in single crystals of silicon-iron by Hull (1960), who showed that the zigzag markings correspond to slip steps on the surface. The markings result if part of the twin is held up below the surface, as illustrated in Fig. 20.16. The slip could arise from the activation of dislocation sources near the end of the stopped twin band, but the results agree better with the suggestion that the slip dislocations are formed by combination of the twinning dislocations at the end of the stopped twin.

When twinning is caused by an external stress, it is clearly necessary that the applied forces do work during the formation of the twin, i.e. that the shear stress across the twinning plane and resolved in the twinning direction should be positive. An important difference between twinning and slip deformation is that twinning is polarized, i.e. reversal of the  $\eta_1$  direction will not produce a twin. This means that, for a single crystal of given orientation with respect to a uniaxially applied stress, some variants of a particular twin mode should operate only in tension, whereas others should operate only in compression. If the single crystal has completely twinned, it follows from Fig. 2.4 that all directions in the initially obtuse sector between  $K_1$  and  $K_2$  will have increased in length, whilst all directions in the acute sector will have decreased. However, Frank and Thompson (1955) pointed out that a slightly different rule will apply to an actual tension or compression test in a single crystal which deforms initially by forming a thin twin lamella. The average shear strain of the specimen is now  $fs$ , where  $s$  is the twinning shear and  $f$  is the volume fraction of the twin. In the limit  $f \rightarrow 0$ , there will be an increase in length for a specimen

axis in the upper right quadrant formed by  $K_1$  and the plane normal to  $\eta_1$  and a decrease in length for an axis in the upper left quadrant.

By analogy with slip, it is natural to enquire whether the onset of twinning coincides with some critical value of the shear stress across the  $K_1$  plane and resolved in the  $\eta_1$  direction. Unfortunately, the scatter in measured twinning stresses is generally too large and the range of orientations available is too small to provide an adequate test of this hypothesis, especially as the incidence of twinning rather than slip as the preferred mode of deformation is itself very sensitive to orientation. It is well established, however, that the initial twin generally corresponds to the variant for which this resolved shear stress is greatest (Allen *et al.*, 1956; Cox *et al.*, 1957; Biggs and Pratt, 1958; Reid *et al.*, 1966). In b.c.c. crystals, for example, with a stress axis within the unit triangle defined by [001], [011] and  $[\bar{1}11]$ , the two variants in tension should be  $(1\bar{1}2)[\bar{1}11]$  near [001] and  $(\bar{1}\bar{1}2)[111]$  near the [001]- $[\bar{1}11]$  boundary, whilst in compression the single variant  $(\bar{2}11)[111]$  should operate over the whole triangle. The different variants imply an asymmetry in the applied stress for twinning in tension and compression and this carries over to polycrystalline textured materials, as shown for iron by Richards and Reid (1968). For materials without a texture, no difference in the twinning stresses for tension and compression was found, but the twinned volume fractions and the fraction of grains containing twins were both appreciably larger in compression than in tension. The authors point out that on average the grains, even in a random aggregate, are more favourably orientated for twinning under an applied compressive stress than under a tensile stress. Similar results apply to f.c.c. metals and alloys where twins form on either the primary or the conjugate {111} slip planes (Narita and Takamura, 1974) or, in the case of constrained deformation, on the most highly stressed system (Chin *et al.*, 1969).

As already indicated, tests on b.c.c. metals give contradictory results in terms of whether or not a critical resolved shear stress law applies, but the majority of the evidence is that the orientation dependence is more complex. Body-centred cubic metals twin readily (without appreciable prior slip) only at low temperatures and/or high strain rates, and under these conditions analysis is complicated by peculiarities in the slip behaviour, which include breakdown of the Schmid law of critical resolved shear stress and large orientation dependencies and asymmetries. In very pure metals, major components of the total strain may result from slip on unpredicted (so-called "anomalous") variants of the usual slip system. These effects are believed to be a consequence of the core structure of the screw dislocation in b.c.c. metals (Christian, 1983; Vitek, 1985) and, later in this section, evidence will be provided that twin nuclei are formed from screw dislocations. Substitutional solutes (especially rhenium) often promote twinning, whereas interstitials may sometimes prevent twinning. It is well established that twinning at low temperatures may often be inhibited by a small prestrain at a higher temperature, but very small prestrains may sometimes increase the number of twins formed upon deformation. A homogeneous distribution of dislocations is most effective in suppressing twinning, and the required prestrain also varies with testing temperature and strain rate (Boucher and Christian, 1972).

Deformation twinning is prevalent in many f.c.c. metals and in alloys of copper, silver or gold, especially at low temperatures, but is not normally encountered in f.c.c. materials of high stacking fault energy such as aluminium and its alloys. The influence of stacking fault energy has been investigated by a number of workers (Thornton and Mitchell, 1962; Peissker, 1965). However, Pond and Garcia-Garcia (1981) have reported that, even in aluminium, twins may form near the tips of cracks in thin foils where the stress concentrations are probably very high.

In normal tensile or compression tests, contrary to the behaviour of many b.c.c. metals, twins form in single crystals of f.c.c. alloys only if prior deformation by slip has so rotated the specimen axis that the primary and conjugate systems are approximately equally stressed.  $K_1$  is then either the primary or conjugate slip plane and  $\eta_1$  is the  $\langle 112 \rangle$  direction of greatest resolved shear stress within that plane. Measurements of this twinning stress, however, do not support a critical resolved shear stress law, but show that twinning becomes easier as the tensile axis approaches  $\langle 111 \rangle$  or the  $\langle 111 \rangle$ – $\langle 110 \rangle$  zone. Narita and Takamura (1974) found that the twinning stress is not very temperature-dependent in Stage II of the stress versus strain curve, but that it decreases as the temperature is lowered if twinning occurs in stage III. Twinning is sometimes accompanied by load drops in f.c.c. alloys but, when the pretwinning slip is small, twinned regions consist of very thin plates and propagate like Luders bands (Fujita and Mori, 1975). The thickness of the plates decreases with decreasing stacking fault energy and temperature, and the onset of twinning is then signalled merely by a bend in the stress versus strain curve.

Twinning is much more common in the h.c.p. metals, largely for reasons of plastic compatibility; the five independent slip systems required for an arbitrary deformation are usually not available. In view of the confusing number of twin modes which have been reported for various h.c.p. metals, it should be re-emphasized here that the important modes are  $\{10\bar{1}2\}$  for all metals and  $\{11\bar{2}1\}$  and  $\{1122\}$  for metals like titanium. The polarization of the shear direction is dependent on the axial ratio and, in the case of  $\{10\bar{1}2\}$  twinning, the shear direction changes sign at an axial ratio of  $3^{1/2}$  which is within the accessible range. For alloys with axial ratios close to the critical value, no twins are formed as the twinning shear tends to zero.

Some investigations on h.c.p. metals (see, e.g., Millard and Thompson, 1952) have given results consistent with the critical resolved shear stress hypothesis, but once again the bulk of the evidence is against such a law. Values obtained by different workers on material of the same nominal purity are widely different, and this suggests, as in some forms of yielding by slip, that any critical stress is extremely sensitive to crystal purity and perfection. In addition, there are convincing reasons for believing that twinning begins from internal stress concentrations which may be produced by prior slip or microslip, and this could lead to considerable variations from sample to sample in the external stress at which twinning takes place. These ideas were strongly supported in some classical experiments by Bell and Cahn (1953) on twinning in zinc. They found that carefully prepared and handled crystals will withstand shear stresses very much higher than that normally required for twinning, but that twinning can be initiated by inducing stress concentrations, for example by pricking with a pin. In these circumstances, it seems clear

that there may be a critical resolved shear stress for growth of a twin nucleus, and that many ordinary crystals may contain such nuclei, but the stress to form a nucleus in a perfect crystal is much higher. The real stress to initiate twinning may be compared to the upper yield point in materials where dislocations are pinned by impurities; the true upper stress is rarely observed because of unavoidable stress concentrations in most mounting and testing procedures. Cahn suggested that twin nuclei should be regarded as tiny stopped elastic twins in Garber's notation.

The importance of imperfections in twin formation was confirmed by experiments on the deformation of highly perfect cadmium and zinc crystals (Price, 1960, 1961, 1962). The specimens, in the form of whiskers or thin platelets, were strained whilst under observation in an electron microscope. The stresses required for nucleation and growth of twins were an order of magnitude higher than those usually measured with macroscopic specimens.

It thus follows that, in discussing the formation of a mechanical twin, its nucleation and subsequent growth must be considered separately, as in most phase transformations. There are, moreover, two distinct growth processes, namely the lengthwise (or edgewise) growth of a lenticular twin constrained in a matrix, and the motion of the habit plane normal to itself, as in a parallel-sided twin crossing a single crystal. If a suitable nucleus exists, or is formed by the applied stress, the first process is usually catastrophic. The twin forms in a very short time, and a "load drop" may be observed in a "hard" testing machine. If the nucleation stress exceeds the growth stress, such load drops may be very large, even in comparison with those caused by sudden yielding in materials like mild steel. The growth velocity must approach the speed of sound, and energy is frequently dissipated as sound waves, giving a characteristic noise known as the twinning cry. The sudden production of a macroscopic twin may be compared to the transition stage from a stopped elastic twin to a residual twin in Garber's terminology, although the twin need not be formed in a single crystal, or extend across the whole cross-section. Many twins of this type form very rapidly (in a few microseconds) and then do not grow or shrink further on increasing or reversing the stress; they have been mechanically stabilized by losing coherency with the matrix as mentioned above. Such twins are sometimes removed by annealing at temperatures where individual atom movements can occur; in other cases, the twins grow at the expense of the matrix on annealing. After removal of the external stress, an assembly containing a constrained twin is in a state of self stress, and the strain energy presumably provides the driving force for the thermal removal or growth of the twin.

The second type of growth may also be very rapid but, under suitable conditions, slow thickening of the twin under an increasing stress can also be observed. The theory of this type of growth in a type I twin with a rational interface is analogous in many ways to that of crystal growth from the vapour. In an ideal crystal with a perfectly planar twin interface, a two-dimensional thermally activated nucleation process would be required to produce growth. Both types of growth also have analogies in the theory of martensitic transformations but, except in special cases, the semi-coherent nature of the martensitic interface makes it easier to envisage the normal growth of a martensite plate than that of the fully coherent twin interface.

In discussing the crystallography of twinning, it was noted that the composition or  $K_1$  plane is the only invariant plane of the lattice deformation relating the two crystals. One way of avoiding high energy surfaces is for the whole twin boundary to consist of planar sections of this type, and a possible model of a tapering or lenticular twin is thus one in which there are occasional steps in the composition plane interfaces. Such steps have already been discussed; they are the twinning dislocations of Section 32. In the case of a lenticular twin, the model contains closed loops of twinning dislocation, parallel to the composition planes, and increasing in radius as the central plane of the lens (Fig. 20.17) is approached.

As discussed in Section 32, twinning dislocations have many of the properties of ordinary dislocations; in particular, they can glide in the composition plane under the action of a shear stress, enlarging or shrinking the twin by one lattice layer as they do so. From Fig. 20.17, we see that a lenticular disc may enlarge its radius simply by expansion of the twinning dislocation loops, the ratio of diameter to thickness becoming greater as it grows in this way. This process will leave increasing areas of flat twin interface near the centre of the twin and, if normal growth also takes place, new twinning dislocations must be formed on these planes. Thus the thickening of a lenticular twin, or the growth of such a twin if the length:thickness ratio is not to increase, must involve the same kind of problems as those involved in the normal growth of a parallel-sided residual twin.

The growth model which has just been described in outline is apparently restricted to compound or type I twins with rational  $K_1$  interfaces and even then there are some complications which arise if  $q$  is greater than two or if the structure has a basis. Suppose first that  $d$  in eqn. (32.1) represents the spacing of the  $K_1$  lattice planes so that the associated Burgers vector  $\mathbf{b}_T$  has magnitude

$$\mathbf{b}_T = d\mathbf{l} \quad (87.1)$$

where  $\mathbf{l}$  is the unit vector parallel to  $\boldsymbol{\eta}_1$ . This step of minimum height is called an elementary twinning dislocation and, as the elastic energy is proportional to the square of the Burgers vector, it might be expected that twinning dislocations with step heights which are multiples of  $d$  will tend to dissociate spontaneously into elementary twinning dislocations. However, when the lattice correspondence does not relate primitive cells of

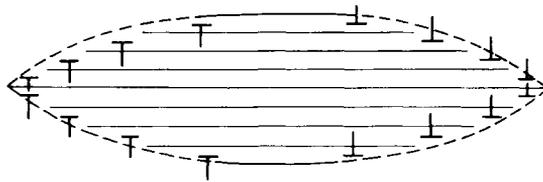


FIG. 20.17. Model of a lenticular twin. If the loops of twinning dislocation are traversed in a constant sense, elements of the loop to the left and right of the figure have the same Burgers vector but opposite line directions. As these elements are in opposition, the dislocation symbols have been reversed. If the dislocations are edges, for example, the long arm of the symbol indicates the compression side of the dislocation.

the two lattices, the two parts of the interface separated by a step of minimum height generally have different structures, and an elementary step may then be energetically unfavourable. The lattice shuffles which accompany all twinning modes with  $q > 2$  imply that the interface structure repeats only at every  $q$  lattice planes parallel to  $K_1$  if  $q$  is odd, or at every  $\frac{1}{2}q$  planes if  $q$  is even, so that the Burgers vector associated with a step between two equivalent regions of coherent interface has a minimum magnitude of

$$\begin{aligned} \mathbf{b}_T &= qds & (q \text{ odd}) \\ \mathbf{b}_T &= \frac{1}{2}qds & (q \text{ even}) \end{aligned} \quad (87.2)$$

This geometrical property of a particular twinning mode was first pointed out by Millard and Thompson (1952) for  $\{10\bar{1}2\}$  twins in h.c.p., and they called the step of minimum height to reproduce the interface structure a "double" twinning dislocation. More generally, such steps are now called "zonal" twinning dislocations (Kronberg, 1959, 1961; Westlake, 1961; Rosenbaum, 1964).

The dissociation of a zonal twinning dislocation into a group of  $q$  or  $\frac{1}{2}q$  elementary twinning dislocations lowers the elastic energy but increases the surface energy. The elementary dislocations have parallel Burgers vectors and hence repel each other until separations are attained at which these repulsive forces are just balanced by the attractive forces due to the excess energies of the various interfaces over that of the minimum energy interface. This situation is clearly analogous to the dissociation of a lattice dislocation into partial dislocations, and the zonal dislocation which is a repeat step of the interface may be regarded as an extended dislocation comprising partial or non-repeat steps separated by "faults", i.e. by regions of higher energy interface. If the energies of these interface faults are sufficiently small, however, the separation of the elementary twinning dislocations will become large and the concept of a zonal twinning dislocation as a separate entity is no longer required.

Zonal twinning dislocations were first discussed in connection with twinning in h.c.p. structures, but the term has also been used to discuss twinning in superlattices in which the minimum step height to give an identical interface structure may be increased to some multiple of that in the disordered structure. The multiple height step is then often referred to as a zonal twinning dislocation, although this is not strictly correct, because  $q$  is now being obtained with reference to the lattice planes of the disordered structure rather than to those of the superlattice. It follows, moreover, from the discussion on pp. 885–887, that the formation of true twins by the motion of such steps would require interchange shuffles at the interface. The confusion arises because in the fully ordered structure there may be atomic planes parallel to  $K_1$  which are equi-spaced and more numerous than the lattice planes. However, in any multiple lattice structure, it is possible to envisage interface steps between atomic planes in which the step height is smaller than that of the elementary twinning dislocation as defined above. Although it would be possible to revise the definition of the elementary twinning dislocation to make it correspond to the minimum step between atomic planes, this would cause difficulties because the twin cannot grow by motion of such a step, and the atomic planes are not necessarily equally spaced. Moreover,

as we shall see below, step defects in which the effective step height is less than the spacing of atomic planes may also arise in some structures by dissociation of elementary twinning dislocations. The definition of an elementary twinning dislocation in terms of the spacing of lattice planes will thus be retained, whilst noting the possibility that step defects of smaller Burgers vector than that given by eqn. (87.1) may be present in some interfaces. Naturally the two interface structures connected by such a step are non-equivalent for any value of  $q$  but, in the special case of superlattice structures, this difference may be confined to changes in the chemical binding, so that the extra energy of the unfavourable interface is similar to that of an antiphase boundary. The rather loose usage to be found in treatments which regard such a step as an elementary twinning dislocation may thus be compared with the way in which extended superlattice dislocations are often described as being dissociated into "lattice" dislocations, i.e. into dislocations with Burgers vectors which are repeat vectors of the *disordered* lattice.

Consider next whether the concepts of step growth can be applied to type II twins, to which the above description does not apply directly because, if  $K_1$  is irrational, the spacing of lattice planes parallel to the interface is indefinitely small. The theory of the atomic structure and of defects in irrational (or very high index) interfaces is difficult and not well developed, and the following description is simplified and partly intuitive. On an atomic scale, an irrational interface must consist of rational facets and, in the general case, a minimum of three differently orientated facets would be required. However, as the interface of a type II twin necessarily contains one rational direction ( $\eta_1$ ), it is possible to model it as alternate facets of two rational planes, as shown in Fig. 20.18. The most appropriate model will generally consist of facets of minimum size parallel to the two closest-packed planes in the zone axis of  $\eta_1$ . If the smaller facets ( $P$ ) all represent transitions between adjacent lattice planes of the larger facets ( $Q$ ), and are spaced along  $Q$  either uniformly or in some regularly repeating sequence at multiples of the distance at which adjacent lattice planes of  $P$  intersect  $Q$ , the average interface will remain rational. However, as the period of the repeat pattern increases, the Miller indices of the interface become higher and higher. The larger facets will be those nearest in orientation to the mean interface plane, and the smaller facets may often be parallel to the rational  $K_2$  plane. These smaller facets may be regarded as steps on the larger facets, and by increasing the period better and better rational approximations to an irrational plane may be obtained. For example, a first approximation to a particular interface might be produced by steps at repeated intervals of five, four and four lattice vectors along the rational  $Q$  interface, and

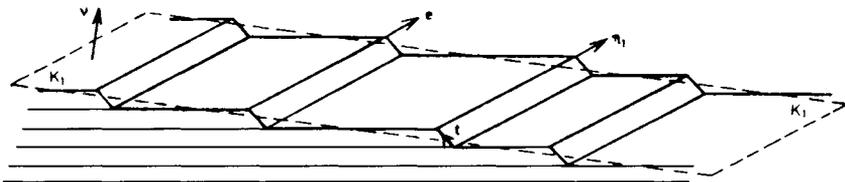


FIG. 20.18. Model of an irrational  $K_1$  interface.

successive approximations might then consist of the following repeated step patterns: ... 544 . . . . . 5444 . . . . . 5444544 . . . . . 54445444544 . . . . . 544454445444544 . . . . . 5444544454445444544 . . . . . etc. In the limit, when the period has become infinite, the interface is truly irrational, but the step pattern remains ordered. The interface is now an example of a one-dimensional quasi-crystal, as there is no genuine periodicity parallel to  $\eta_1$  but the structure is quasi-crystalline (Levine and Steinhardt, 1986) in the interface direction normal to  $\eta_1$ . (The general theory of quasi-crystals became significant after the discovery of three-dimensional quasi-crystals with fivefold symmetry [see pp. 617]). Sutton (1988) has shown that the "structural unit" model of irrational tilt boundaries with a rational tilt axis (Sutton and Vitek, 1983) may be treated as a one-dimensional quasi-crystal in which the periodic combination of *A* and *B* structural units becomes quasi-periodic in the limit. The theory has also been applied to "hidden symmetries" in generalized grain boundaries by Gratias and Thalal (1988).

Note that if the plane of the larger facets were parallel to the true  $K_1$  plane, the steps would be twinning dislocations in screw orientation. For a matching irrational plane, however, they are an essential element in the structure of the boundary, and have been called "intrinsic twinning dislocations" (Guyoncourt and Crocker, 1968). If an irrational  $K_1$  interface traverses a single crystal, there will be no force driving the steps in a particular direction when a stress is applied to the crystal unless a nucleating mechanism is available to supply fresh steps as required to maintain the irrationality of the interface plane. If this can be achieved, however, the steps will glide along the rational planes as the interface moves forward, and it is in this sense that they can also be regarded as twinning dislocations. This leads to the concept that steps on the rational facets could also assume edge or general orientations and local densities other than those specified by the lattice matching condition. The excess steps over the ideal model of the interface now constitute extrinsic twinning dislocations which will respond to an external stress in the same way as steps in rational  $K_1$  interfaces. Extrinsic twinning dislocations are effectively steps in the macroscopic irrational interface, so that their motion displaces this interface.

The magnitude of the effective Burgers vector of such a twinning dislocation is given by the twinning shear multiplied by the height of the step in the irrational plane, i.e. by the interplanar spacing of the  $Q$  lattice planes resolved in the direction of the normal to the irrational  $K_1$  plane. Hence the Burgers vector is

$$\mathbf{b}_T = u_i m_i s \mathbf{l} = \bar{d} s \mathbf{l} \quad (87.3)$$

where  $\mathbf{u}$  is any vector connecting lattice points on adjacent planes of type  $Q$ , and  $\bar{d}$  is the spacing of the  $Q$  planes resolved in the direction of the unit normal  $\mathbf{m}$  to the irrational interface. Note that the magnitude is irrational, but that the vector is parallel to the rational  $\eta_1$  direction. This step in plane  $Q$  is the type II equivalent of an elementary twinning dislocation, but it follows from the discussion on p. 916 that the interface structure on the two sides will be identical only if a primitive lattice vector  $\mathbf{w}$  parallel to  $\eta_1$  traverses not more than two lattice  $K_2$  planes. In other cases, we must define the equivalent

of the zonal twinning dislocation by

$$\mathbf{b}_T = w_i m_i \mathbf{s}_i = \bar{q} \bar{d} \mathbf{s}_i \quad (87.4)$$

It might be considered that the intrinsic steps of the irrational interface would all have a height of  $\bar{q} \bar{d}$  rather than  $\bar{d}$ , but this cannot be assumed without modelling the actual interface structure; it is conceivable, for example, that an alternating configuration at the  $Q$  facets is more favourable than a repeated configuration.

An equivalent formal treatment of a faceted interface is to use the Frank–Bilby–Bollmann equation (see Chapter 8) to calculate separately the formal dislocation content required to correct the misfit along each of the facets  $P$  and  $Q$ . This corresponds to  $-\mathbf{b}_T$  for an individual step facet  $P$ , but sums to zero over a sufficient number of facets if the relative lengths of  $Q$  and  $P$  perpendicular to  $\boldsymbol{\eta}_1$  give the exact  $K_1$  interface. Various ambiguities in the dislocation descriptions of stepped and unstepped interfaces depend on how zero dislocation content is defined, and are discussed by Olson and Cohen (1979), Olson (1981), Christian and Knowles (1982) and Christian (1983).

Returning to rational twin interfaces, consider the twinning dislocations in the important metallic structures. Calculations indicate that in some twin boundaries there may exist alternative interface configurations of nearly equal energies, and the line of separation of two such regions may then be a dislocation with a Burgers vector smaller than that of an elementary twinning dislocation and a step height equal to a fraction of the spacing of atomic  $K_1$  planes. This applies specifically to the coherent  $\{112\}$  interface of a b.c.c. twin for which the two possible structures have been revealed by computer calculation (see p. 353). It was originally concluded that the configuration shown in Fig. 8.12(b) will have the lowest energy, except possibly for very “soft” atoms which have a pairwise potential in which the repulsive part increases relatively slowly with decreasing separation. Later work, however, indicates that the computed equilibrium structure is sensitive to the boundary conditions and method of relaxation as well as to the potential, so that it is not easy to predict what is the equilibrium interface structure for any real material. Some calculations (Bristowe and Crocker, 1975; Yamaguchi and Vitek, 1976) indicate that the lowest energy structure may not be the same in all b.c.c. metals, and also that the difference in the energies of the two structures may be very small in some materials. Bristowe and Crocker pointed out that this leads to the possible dissociation of an elementary twinning dislocation into two partial twinning dislocations separated by a strip of interface in the configuration which has the slightly higher energy.

Figure 20.19 includes the same two configurations of the b.c.c.  $\{112\}$  twin boundary as Fig. 8.12. By making different atomic connections, the displaced boundary (f) is seen to consist of a layer of cells which are represented in projection as interlocking isosceles triangles. Bristowe and Crocker refer to the two structures as reflection and isosceles twin boundaries respectively; note that the planar geometrical interface is coincident with a plane of atoms for the reflection (or coincidence site lattice) boundary but is midway between two atomic planes for the isosceles boundary. This diagram illustrates more clearly than did Fig. 8.12 that both boundaries appear identical when viewed from either side of the interface.

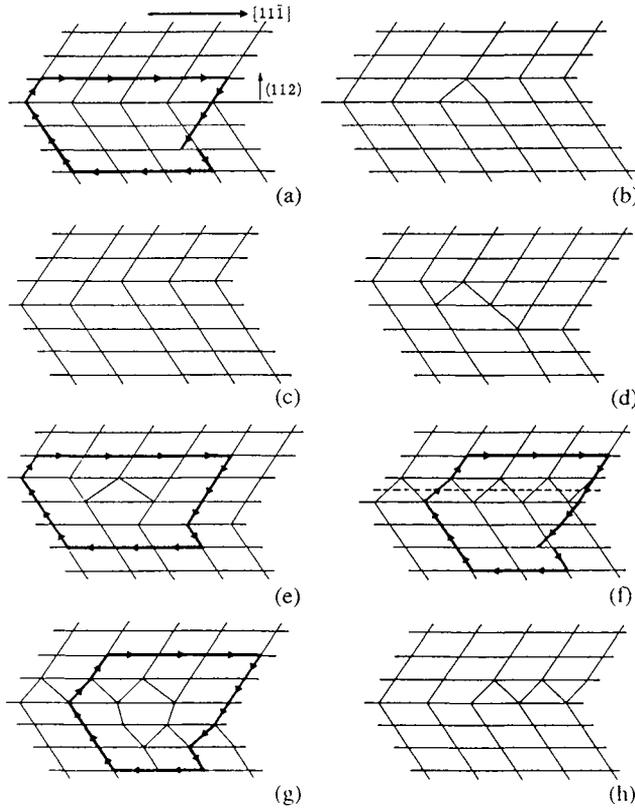


FIG. 20.19. The structures of possible twinning dislocations in b.c.c. crystals projected onto a  $(\bar{1}\bar{1}0)$  plane which has a twofold stacking sequence. A perfect reflection twin boundary is shown at (a) and this boundary contains perfect complementary, double and zonal twinning dislocations at (b), (c), (d) and (e). A perfect isosceles boundary is shown at (f) and this boundary contains an elementary dislocation at (g). A partial twinning dislocation is shown at (h).

An elementary twinning dislocation in a reflection type boundary is shown schematically in Fig. 20.19(b) and that in an isosceles boundary is shown in Fig. 20.19(g); Fig. 20.19(h) shows a reflection and an isosceles section of interface separated by a partial twinning dislocation of step height equal to one-half of the  $(112)$  plane spacing and a corresponding Burgers vector which is one-half that of the twinning dislocations of Fig. 20.19(b) and (g). In contrast to most extended lattice dislocations, there is no orientation of an applied shear stress which will tend to drive two partials from a dissociated elementary dislocation apart, as their Burgers vectors are parallel.

When a zonal twinning dislocation (or an elementary twinning dislocation if the twin mode has  $q < 3$ ) is displaced along the  $K_1$  interface, there is no first-order change in energy, other than the self energy of the dislocation, which may oscillate with the periodicity of the lattice in the direction of displacement. The dislocation may thus be

described as glissile, and its motion provides a mechanism for the growth or shrinkage of the twin, as described above. The lattice resistance to motion may be regarded as a friction stress arising from a kind of Peierls–Nabarro force (see pp. 273–280), and it depends on the detailed atomic structure of the step or dislocation core. In particular, if the core extends over several atomic distances along the  $K_1$  plane, the lattice frictional resistance may be expected to be relatively small.

In early treatments of the possible nucleation and growth processes leading to the formation and subsequent spreading of new layers of the twin orientation, it was generally assumed that the core of a twinning dislocation is similar to that of a lattice dislocation and is thus quite narrow. Some justification for this assumption arises from the magnitudes of the Burgers vectors in cubic metals and especially from the close similarity between the elementary twinning dislocation in f.c.c. metals and the Shockley partial. Experimental evidence that small steps in twins are fairly sharp discontinuities has been provided by direct observation of elementary twinning dislocations which are visible when foils of f.c.c. copper (Mahajan *et al.*, 1970) or b.c.c. molybdenum–rhenium alloys (Mahajan, 1972a) containing very thin tapering twins are examined in the transmission electron microscope. However, computer calculations by Yamaguchi and Vitek (1976) and by Bristowe and Crocker (1977) indicate that twinning dislocations, at least in b.c.c. structures, are appreciably wider than lattice dislocations, and this may have important implications for the theory of twin growth. The calculations of Bristowe and Crocker (1976, 1977) have also revealed the possible existence of other step defects in b.c.c. twin boundaries, in addition to the elementary and partial twinning dislocations already described. These defects are shown in Fig. 20.19(c) and (e) and are called complementary and zonal twinning dislocations respectively.

Figure 20.19(c) shows the complementary twinning dislocation which was originally introduced by Sleswyk (1962). It has the same step height as the elementary twinning dislocation [Fig. 20.19(b)] but, for the same sense of the step, its Burgers vector is twice as large and in the opposite direction, so that its structure is quite different. Sleswyk introduced this defect by considering the hypothetical dissociation of a twinning dislocation in a tapering twin into a lattice dislocation and a complementary twinning dislocation according to the Burgers vector equation

$$a/6\langle 111 \rangle = a/2\langle 111 \rangle + a/3\langle \bar{1}\bar{1}\bar{1} \rangle \quad (87.5)$$

The reverse process of combination of a twinning dislocation with a lattice dislocation of antiparallel Burgers vector is clearly an equal formal possibility; in one case, the lattice dislocation is emitted from the interface, and in the other case the opposite lattice dislocation glides into the interface. In fact, however, with the pairwise potentials that were used by Bristowe and Crocker, an isolated complementary dislocation is unstable and will dissociate into elementary, partial and zonal twinning dislocations. Motion of a complementary dislocation along the  $K_1$  plane is thus an unlikely process, but we should note that if it were to occur, the associated shear would be  $2^{1/2}$  and if successive (112) planes were displaced in this way, the twinning mode produced (see Table XVI) would have  $K_1 = \{112\}$ ,  $K_2 = \{110\}$ ,  $\eta_1 = \langle 11\bar{1} \rangle$  and  $\eta_2 = \langle 001 \rangle$ .

A step defect with the same Burgers vector as that of the complementary twinning dislocation but a double step height in the opposite sense is shown in Fig. 20.19(d). This simply consists of two coincident elementary twinning dislocations [each of opposite sense to that shown in (b)] and its motion produces the usual twinning shear of  $2^{-1/2}$ ; it is, of course, unstable against dissociation into its component twinning dislocations. Consider now, however, the combination of a double step of this kind with a slip dislocation, as represented on the right-hand side of eqn. (87.3), to form a step defect of double step height but with a Burgers vector of type  $a/6\langle 111 \rangle$ . This is the zonal twinning dislocation shown in Fig. 20.19(e) for a reflection type boundary.

The existence of a zonal twinning dislocation in a  $\{112\}$  b.c.c. twin interface was unpredicted prior to the computer simulation, and it should be emphasized that this is not a zonal twinning dislocation of the usual b.c.c. twin mode. As the Burgers vector has the same magnitude as that of the usual twinning dislocation but the step height is twice as large, it follows that propagation of such a dislocation through a series of  $\{112\}$  planes will produce a shear of  $2^{1/2}/4$ . The corresponding twinning mode has  $K_1 = \{112\}$ ,  $K_2 = \{3\bar{3}\bar{2}\}$ ,  $\eta_1 = \langle 11\bar{1} \rangle$ ,  $\eta_2 = \langle 113 \rangle$  and  $q = 4$ , so that one-half of the lattice points (atoms) must shuffle as the interface moves forward. (This mode is listed in Table XVI.) The defect shown in Fig. 20.19(e) is thus correctly described as a zonal twinning dislocation of this hypothetical, smaller shear twinning mode, but the usage is rather loose when the defect is present in the boundary of a deformation twin which has formed by a different shear, or that of a non-deformation twin.

It is useful to digress here to consider the general question of steps and line defects in coherent twin boundaries. The Burgers vector specified by eqn. (32.1) is unambiguous for a given step height, but the equation is only relevant to the growth mechanism if the step has formed by a physical shear which is equivalent to the virtual process of Fig. 7.16. It is clear that, in discussing the possible defects which can arise in a coherent interface, only the positions of the atoms on the two sides of the interface are significant, and the origin of the interface is relevant only if consideration is limited to defects which may have formed in a particular way. In this sense, the possible Burgers vectors of step defects in coherent annealing twin boundaries may be found, for example, by supposing the twin and its step to have first formed by the virtual process of Fig. 7.16, followed by any rearrangement of the structure which does not remove the stepped interface. It is clear that, although such rearrangements will lead to various possible effective Burgers vectors for a step of given height, it will not generally be possible to remove the elastic discontinuity completely and so produce a pure step.

Some authors (Balluffi *et al.*, 1972; Ashby, 1972) have distinguished between grain boundary dislocations and grain boundary steps for the general grain boundary, and have suggested that there is no strain field associated with an ideal step. However, it needs to be emphasized that formation of an ideal step is rarely possible even when the step height is a multiple of the interplanar spacing; a necessary condition (Christian and Crocker, 1980) is that the plane defined by the step length and the riser is a lattice plane of a coincidence site lattice of the two crystal lattices. In the cases of coherent twin boundaries on  $\{111\}$  planes of f.c.c. crystals or  $\{112\}$  planes of b.c.c.

crystals, pure steps are possible when the step height is some multiple of three interplanar spacings.

It should now be clear that, if no restrictions are placed on the mechanism of formation of an interface of given structure, the set of possible Burgers vectors associated with a step of fixed height in this interface is obtained by taking any representative Burgers vector of the set [given, for example, by eqn. (32.1) or by a DSC lattice construction] and adding to it any lattice repeat vector. This procedure may be compared with that used in Section 33 to derive the possible Burgers vectors of the imperfect dislocations which may bound a given stacking fault. It should also be emphasized that, although it is possible to use different virtual lattice deformations to define the same interface, this does not lead to any ambiguities in the possible Burgers vectors of a given step; these vectors form a unique set and may be derived from any representation of the lattice relations. Pure steps may be envisaged if, and only if, this set contains a zero Burgers vector.

It is clear that the complete set of Burgers vectors of "perfect" interface dislocations (i.e. linear defects which separate two regions of identical interface structure) is identical with the set of vectors which constitute the DSC lattice of Bollmann (1970), as already discussed on pp. 373–375. These vectors link the sites of the parent lattice to those of the twin when the two lattices are in a coincident orientation and position, i.e. when  $\mathbf{t}=0$ . For a single lattice structure, the DSC dislocations are the only vector translations which reproduce the interface structure. The vectors are independent of  $\mathbf{t}$  (see p. 374) and each may be associated with a step of height  $d$ . However, in a general discussion of possible defects which reproduce the structure of an interface, Pond (1985, 1995) has shown that, in non-symmorphic crystals (i.e. crystals containing mirror glide planes or screw rotation axes), additional dislocation type defects may in certain circumstances exist in an interface and reproduce the interface structure. Pond's theory follows on from his general treatment (Pond and Bollmann, 1979; Pond and Vlachavas, 1983) of "dichromatic colour" (or "black and white") symmetry in interface structures mentioned briefly on p. 375. He shows that, whereas defects in single crystals are characterized by the symmetry operations of translation, proper rotation or proper screw-rotation, giving rise respectively to dislocations, "disclinations" (see pp. 257–258) and "dispirations" (Harris, 1970), interface defects are characterized by combinations of symmetry operations, one from each crystal.

In Pond's formulation, the twinning dislocations considered above arise from the broken translational symmetry of the interface but, in some interfaces in non-symmorphic crystals, it is also possible to produce interface defects of translational character by combinations of point symmetry operations which are aligned but contain intrinsic glide components which are either not equal or not parallel. These "supplementary displacement dislocations", unlike the DSC dislocations, have Burgers vectors which are modified by any displacement  $\mathbf{t}$  which breaks the translational alignment of the symmetry elements of the two crystals. Moreover, in general, such defects arise only at the junction between two differently orientated (but equivalent) interfaces and they only represent defects in a single interface when special conditions have been satisfied. The general theory of such defects (Pond, 1995) is too complex to be included here, but some special cases in connection with twinning dislocations in h.c.p. crystals will be considered.

In general, the atomic configuration at a step may be expected to adjust itself wherever possible so as to minimize the effective Burgers vector. This implies that the structures of minimum height steps in corresponding boundaries of annealing and deformation twins are likely to be identical, as the operative twinning mode of the deformation twin is likely to correspond to a minimum shear and hence to a minimum Burgers vector condition. However, in an annealing twin, there is likely to be an equal density of steps with each of the possible crystallographically equivalent directions of the Burgers vector which lie in the coherent interface (three for f.c.c. twinning) whereas, in the deformation twin, the steps will predominantly have Burgers vectors parallel to the physical  $\eta_1$  direction.

For a f.c.c. twin, the structure of the fully coherent  $\{111\}$  interface is expected to be little relaxed from the ideal structure produced by shear without any relative translation of the two lattices. This follows from a pairwise model because, as pointed out on p. 121, differences between the various structures corresponding to ideal close-packing of spheres only appear when the computations are extended to at least third nearest neighbours. Thus the energies of an intrinsic stacking fault in the f.c.c. structure, the  $\{111\}$  twin boundary and the  $\{111\}_c//\{00.1\}_h$  interface between the f.c.c. and h.c.p. structures are all likely to be similar and small (see pp. 122–124), and this is confirmed by the calculations of Vitek (1968) on stacking faults and those of Sutton and Christian (1982) for the cubic–hexagonal interface. The elementary twinning dislocation for f.c.c. has the same Burgers vector as a Shockley partial and a step height equal to the spacing of the  $\{111\}$  planes. As in the b.c.c. case, defects with other Burgers vectors of the DSC lattice are theoretically possible; for example, with the same step height and a Burgers vector in the antitwinning direction of twice the magnitude, corresponding to an elementary twinning dislocation of the 1.3 mode of Table XVII. However, these other type defects, except possibly pure steps with a height of three planes, or some integral multiple of three, seem unlikely to occur in practice. As there is probably only one stable configuration of the interface, there are unlikely to be any partial twinning dislocations in this structure.

Equivalent defects cannot form in h.c.p. twin boundaries as the basal planes of this structure are the only close-packed planes and cannot act as  $K_1$  planes because they are already mirror planes. Simulations of the interface structures of the main h.c.p.  $K_1$  interfaces, namely  $\{10\bar{1}2\}$ ,  $\{11\bar{2}1\}$  and  $\{11\bar{2}2\}$ , have been made by Serra and Bacon (1986) using three different two-body potentials. Only one equilibrium configuration was found for the  $\{10\bar{1}2\}$  interface, and the parent and twin structures are mirror images in the interface. The plane of the interface is an atomic plane formed by the coalescence of two adjacent original atomic planes which make up a corrugated lattice plane of type  $\{10\bar{1}2\}$ . This is the symmetrical configuration anticipated from Fig. 20.11, and discussed on p. 897, and the orientation relation is both type I and type II (i.e. “combined”) on an atomic scale. It follows that the only stable twinning dislocation expected for this mode is the zonal dislocation of double step height first discussed by Millard and Thompson. This has a Burgers vector (in the four-axis basis  $a, a, a, c$ )

$$\mathbf{b}_T = [(3 - \gamma^2)/(3 + \gamma^2)](10\bar{1}1) \quad (87.6)$$

so that

$$|\mathbf{b}_T| = [(3 - \gamma^2)/(3 + \gamma^2)^{1/2}]a \quad (87.7)$$

The Burgers vector thus has magnitude  $a/(51)^{1/2} \cong a/7$  for ideal axial ratio,  $\gamma = (8/3)^{1/2}$ , but varies rapidly with  $\gamma$ . Computer simulation of the structure using two-body potentials shows that the width of the twinning dislocation is sensitive to the assumed potential.

The stable interface of a  $\{11\bar{2}2\}$  twin is also an atomic plane of reflection according to the calculations of Serra and Bacon but, because the motif unit lies in the  $\{11\bar{2}2\}$  planes, all the lattice planes parallel to the interface are flat in this case. The interface structure is thus that of a type I twin, but also possesses the additional symmetry discussed on p. 903. This structure was the only stable interface found by Serra and Bacon for so-called equilibrium pair potentials but, when non-equilibrium potentials were used, i.e. when the crystal was subjected to an external pressure which corresponded roughly to the Cauchy pressure, an additional stable interface was formed by relative translation of the two sets of atoms by  $\mathbf{t} = [1/6(1 + \gamma^2)] \langle 11\bar{2}\bar{3} \rangle \cong (1/22)\langle 11\bar{2}3 \rangle$  parallel to the  $\boldsymbol{\eta}_1$  direction of the  $g = 6$  twinning mode. This is the translation required to obtain a type II interface, but the twofold axis is midway between two atomic  $K_1$  planes. This atomic symmetry might, however, be built into the pair potential model as the potentials used by Serra and Bacon do not well represent real materials. One of the non-equilibrium potentials also gave a third possible and less symmetrical atomic model of the interface produced by an alternative displacement parallel to  $\boldsymbol{\eta}_1$ , but this was not stable when the other potentials were used.

As already noted, the probable twinning mode has  $g = 6$ , so that the important twinning dislocation will be expected to be zonal, with a step height equal to three interplanar spacings of the  $\{11\bar{2}2\}$  lattice planes. The core structure of this dislocation will depend on the details of the shuffle displacements, and Serra *et al.* (1988) made computations for three different possibilities. The energy and width of the step were found to be not very sensitive to the shuffle model used, but to depend on the potential. Calculations were also made for two other possible step defects in this type of boundary, having step heights of one and four interplanar spacings respectively. These steps would be twinning dislocations of other hypothetical twinning modes for which  $K_1 = \{11\bar{2}2\}$ ; in the single step case, the mode has a high shear of  $\sim 1.22$  and the four-layer step has more complex shuffles. Thus these dislocations are probably not important in the mechanism of deformation twinning.

Two different starting configurations were considered by Minonishi *et al.* (1982) for the atomic structure of a  $\{11\bar{2}1\}$  interface, and these were called *R* and *D* respectively. Using a Lennard-Jones potential, they found two correspondingly stable relaxed structures for the interface, neither of which corresponds to mirror symmetry on an atomic scale. Similar results were obtained by Serra and Bacon, who pointed out that the lower energy (relaxed *D*) interface may be obtained from the mirror interface (unrelaxed *R*) by moving alternate basal planes in opposite senses through  $(1/12)\langle 1\bar{1}00 \rangle$  to change the stacking from  $\dots ABAB\dots$  to  $\dots ACAC\dots$ . The configuration of the higher energy (relaxed *R*) interface is similar, but it has a row of vacancies in each *B*-*C* transition region. The relaxed

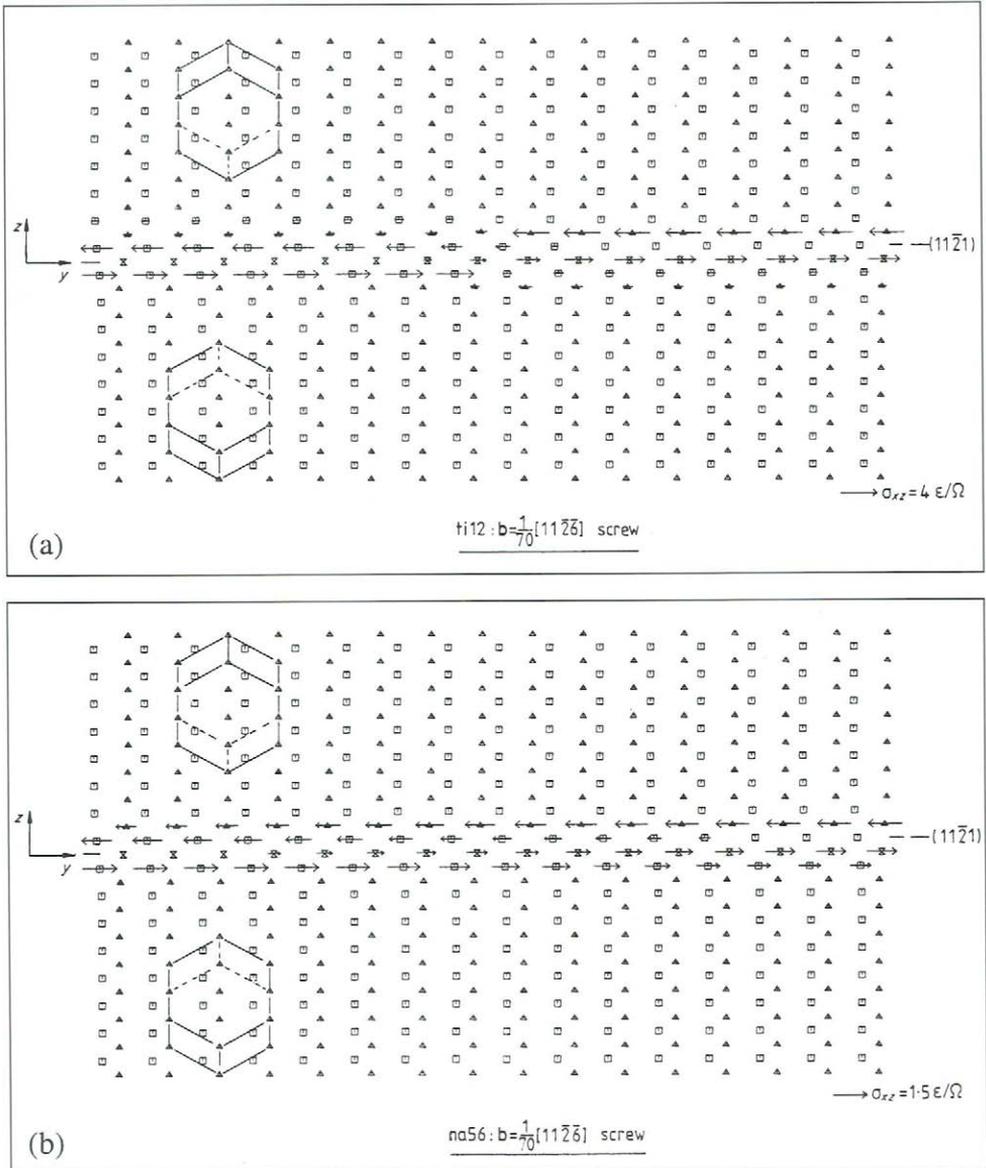


FIG. 20.20. A twinning dislocation in a  $\{11\bar{2}1\}$  interface computed with two different pair potentials (after Serra *et al.*, 1988).

$D$  interface is shown in Fig. 20.20; it again has an atomic interface plane, but in this case it is an actual atomic plane of each crystal, not a coalesced lattice plane. Serra and Bacon point out that the atom positions are connected by a twofold axis along  $\eta_1$  and lying at level 0 in the atomic interface of Fig. 20.20 although this symmetry was not present in the

starting configuration. As no lattice shuffles are involved in  $\{11\bar{2}1\}$  twinning an elementary twinning dislocation reproduces the interface structure. This dislocation is specified by

$$\mathbf{b}_T = [1/3(4\gamma^2 + 1)](11\bar{2}\bar{6}) \quad (87.8)$$

and

$$|\mathbf{b}_T| = [2/(3 + \gamma^2)]^{1/2} a \quad (87.9)$$

For ideal axial ratios, the Burgers vector is  $\mathbf{b}_T \cong (1/35)(11\bar{2}\bar{6})$ . When this dislocation was simulated, however, Serra *et al.* found that it decomposed into two dislocations, each with one-half of the above Burgers vector, as first envisaged by Minonishi *et al.* These two dislocations each have a step height equal to the separation of atomic planes parallel to the interface, i.e. to one-half of the spacing of lattice  $K_1$  planes. Although this ordinarily would not reproduce the interface structure, the non-alignment of the translational components of the mirror glide symmetry of the  $\{1\bar{1}00\}$  plane of shear ensures that the two interfaces bounding such a step have equivalent structures and energies. This is an example of the "supplementary displacement dislocations" mentioned on p. 923. With two of the three potentials used, Serra *et al.* found the core width of both edge and screw twinning dislocations of this type to be very wide, implying that the steps may be very mobile.

As already noted, the edge of a twin wedge or lenticular plate in a stressed material will be a region of high stress concentration, and may be compared with a crack tip or an incomplete kink band. If the yield stress of the matrix is sufficiently high, an enclosed twin may be accommodated elastically and an adequate treatment of its stress field and of the energy of the assembly may be obtained by Eshelby's method (see pp. 466-471), which gives analytical solutions in the isotropic approximation for a twin of assumed ellipsoidal shape. Such a calculation is relevant to the possible homogeneous nucleation of a twin within a grain or in a region of stress concentration, as proposed by Orowan (1954) and discussed below. The high stress at the tip of an elastic twin will tend to produce further growth of that twin, as was shown for the analogous case of an incomplete kink band in a classical treatment by Frank and Stroh (1952).

If the twin is not completely elastic, i.e. if there has been some plastic accommodation as shown in Fig. 20.15, it becomes much more difficult to calculate the stress field and the energy, just as it does for an elastic-plastic crack. Thus it becomes important to discuss the problem of accommodation, already mentioned in connection with twin morphology, and the experimental measurement of the twinning shear in more detail. Before doing so, however, it should be mentioned that models which are not purely elastic may be used to calculate the stress field and energy of an enclosed twin. It may be regarded, for example, as a continuous distribution of infinitesimal dislocations, or attempts may be made to derive an atomistic model with some assumed interatomic potentials. A brief review of all three methods has been given by Sree Harsha (1981).

Accommodation slip at the ends of an internally terminating b.c.c. twin generally takes the form of a single slip band parallel to the  $K_1$  plane although, in silicon-iron crystals, slip sometimes occurs also on those intersecting  $\{110\}$  type planes which contain the Burgers

vector of the twinning dislocation. The slip dislocations may form directly from the twinning dislocations of the interface, as originally proposed by Sleswyk (1962), or may be nucleated ahead of the interface.

Sleswyk considered a dissociation of the type represented in eqn. (87.5) and supposed that this occurs in every third twinning dislocation in a tapering boundary. The lattice dislocations are then assumed to move away on their glide planes, which are of course planes parallel to the  $K_1$  plane of the twin, and they leave behind a boundary consisting of groups of two elementary twinning dislocations and one complementary twinning dislocation. If the sequence is repeated nearly regularly, the net Burgers vector content of the new twin boundary is zero, so that its motion produces no net shear but nevertheless converts the parent structure into the twin structure. In contrast, the shape shear formerly associated with the twin is now produced in a region ahead of the twin by the glide of the regularly spaced lattice dislocations. The new twin boundary may be assumed to have a lower energy than the original boundary, but only because additional energy has been supplied to create the lattice dislocations; the dissociation in eqn. (87.5) is probably energetically very unfavourable. The postulated slip dislocations are supposed to have been emitted from the original twin interface and were called emissary dislocations by Sleswyk.

It has been proposed that, instead of the dissociation reaction, loops of slip dislocation are created spontaneously in the high stress field near the tip of the twin. Each such loop will then expand, part of it interacting with the twin boundary to produce an  $a/3[111]$  complementary dislocation in the interface, and the remainder spreading away from the twin. Clearly, if loops are nucleated on every third lattice plane, the final result will be indistinguishable from Sleswyk's emissary slip.

The theory of emissary dislocations implies that the shape deformation associated with twin formation may be physically separated from the actual change of lattice orientation. The suggested model of the boundary and its associated emissary dislocations is shown in Fig. 20.21; the instability of an isolated complementary dislocation predicted by Bristowe and Crocker will not necessarily apply to the regularly spaced complementary dislocation of this array, which has no long-range field. In any event, if a set of emissary dislocations can be formed, the twin boundary will have some structure which produces no net shape change so that, if this boundary advances into the region already swept by the emissary dislocations, the total change is indistinguishable from the usual picture of twinning. Direct evidence for emissary slip ahead of a stopped twin in a molybdenum-rhenium alloy has been obtained by Mahajan (1975). The dislocations ahead of the twin tip were confined to a slip zone coplanar with the twin, as expected, but were of both signs, so that some modifications have to be made to the simple model.

A growing twin might have one of the two general shapes shown in Fig. 20.22 and electron microscopic observations have revealed configurations of these types (see, e.g., Votava and Sleswyk, 1962). In some cases, parts of the twin interface may be held up by obstacles so that irregular shapes which differ appreciably from the ideal picture of Fig. 20.22 are produced. This is illustrated in Fig. 20.23. It will be seen that the separation of the lattice and shape changes allows the twin to occupy only part of the region

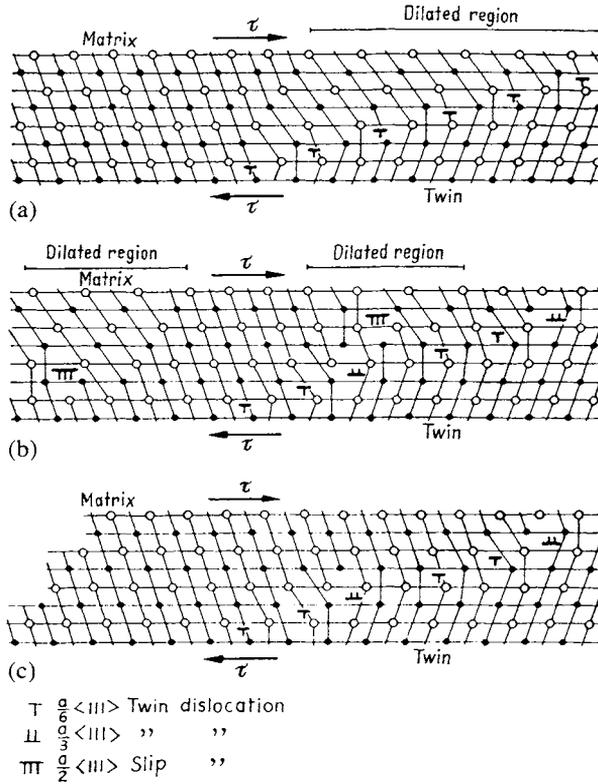


FIG. 20.21. The dissociation of a twin termination boundary in the b.c.c. structure (after Sleeswyk, 1962). Two layers of atoms are projected onto a (110) plane. In (a) the boundary between twin and matrix is represented as an array of  $1/6\langle 111 \rangle$  twinning dislocations on successive  $\{111\}$  planes. In (b) the dissociation has produced the emissary dislocations which move away from the interface under the influence of the stress, and complementary  $1/3\langle 111 \rangle$  type twinning dislocations. The residual boundary shown in (c) has one complementary dislocation to every two twinning dislocations and produces no far-reaching strains in the lattice.

of homogeneous shear produced by the emissary dislocations, and thus to have an irregular shape.

An alternative view is that there are no emissary dislocations ahead of a moving twin interface but that, when the interface is halted for some reason, emissary slip occurs to ensure continuity of the shear and hence compatibility of the shape deformation. This would still allow the formation of twins of irregular shape, as not infrequently observed, and would avoid the difficulty that there is no driving force tending to displace the twin interface of Fig. 20.21(c).

Nearly equivalent results could be obtained in principle if the stresses at the edge of a twin merely activated dislocation sources, but the slip would then not be so homogeneous, nor confined in general to the twin plane. Even in silicon-iron, where slip on  $\{110\}$  planes

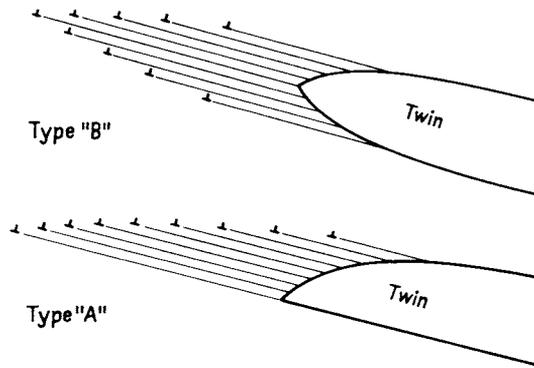


FIG. 20.22. Diagram illustrating two types of emissary dislocation array (after Votava and Sleeswyk, 1962). Associated twin fronts have been observed in molybdenum rhenium alloys.

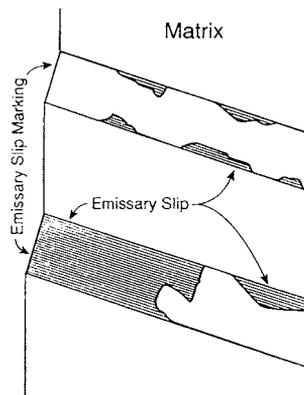


FIG. 20.23. Schematic drawing to show that twins need occupy only part of the regions of emissary slip (after Sleeswyk, 1962).

is observed to emanate from the tip or from irregularities in the side faces, there are no apparent pre-existing lattice dislocations to act as sources, and Hull suggested that, in such b.c.c. materials,  $(1/2)a(111)$  slip dislocations are nucleated at or near the interface but, in screw orientation, cross-slip on to  $\{110\}$  planes as they move away from the region of intense stress concentration. Very approximate calculations suggest that the energetic conditions favour emissary dislocations when the thickness of the twin lamellae exceeds about 4 nm at the normal stresses at which twinning occurs in iron.

Emissary slip of the above type cannot be expected in other crystal structures as the Burgers vectors of twinning and slip dislocations are not normally parallel. Nevertheless, whenever a twin is stopped within a surrounding matrix, it should generally be possible to accommodate (i.e. reduce) the shear discontinuity at the tip by conversion of some of the twinning dislocations into "emissary" lattice dislocations. Thus, Pond and Garcia-Garcia

(1981) observed dislocation structures near twin tips in thin foils of aluminium (see p. 913) which were consistent with the formation of two "residual" Shockley partials and one "emissary" f.c.c. slip dislocation from three adjacent twinning dislocations.

Now consider the problem of growth normal to a rational  $K_1$  plane. The twin might extend by one repeat distance ( $qd$  or  $\frac{1}{2}qd$ ) by first forming a small island or two-dimensional nucleus, which could then spread outward under the applied stress, but a very small island will tend to collapse again because of the line tension of the peripheral twinning dislocation. A region sufficiently large to expand continuously may thus be established only by a favourable series of thermal fluctuations, so that in this idealized model there may be a nucleation difficulty at each successive equivalent position of the twin interface. Of course, there may be favoured regions of high internal stresses where steps in the interface can be more readily nucleated, for example the intersections of the coherent twin boundary with a grain boundary or a free surface.

The nucleation of a closed loop of twinning dislocation by the combined effects of thermal agitation and applied stress may be compared to the corresponding problem of the spontaneous creation of a lattice dislocation, discussed on p. 281. For a loop of given radius, the activation energy according to a linear elastic model is proportional to the square of the Burgers vector [eqn. (31.17)], and the stress required is proportional to the Burgers vector [eqn. (31.16)]. Thus in structures with a sufficiently small twinning shear, it is possible that normal growth at finite temperatures does involve spontaneous nucleation of little islands of twin on each successive lattice plane. For example, if the Burgers vector of the twinning dislocation is about one-tenth of the interatomic distance, the assumptions made on p. 281 imply that, with a stress of  $\sim 8 \times 10^{-4} \mu$ , the activation energy is only  $\sim 0.2 \text{ eV}$  and the critical radius is  $\sim 10 \text{ nm}$ . However, if the twinning dislocation has a Burgers vector greater than (say) one-third of the interatomic distance, as in cubic crystals, the probability of spontaneous nucleation will (with a linear elastic model) be almost negligible, as it is for perfect lattice dislocations.

Twinning dislocations in face-centred tetragonal or orthorhombic structures, such as those found in indium thallium or gold-cadmium alloys, have Burgers vectors with magnitudes in the range 0.006–0.04 nm. Birnbaum and Read (1960) used eqns. (31.16) and (31.17) to show that twin boundary motion by spontaneous nucleation is quite probable in such alloys, the required activation energy being comparable to the available thermal energy at stresses corresponding to those at which the boundaries are observed to move. In these alloys, a single crystal of the high temperature cubic phase may be transformed martensitically to give a specimen consisting of a single set of parallel twins (see p. 1030). The twins are extremely mobile under small stresses and deform plastically entirely by movement of the twin boundaries, up to a limiting strain. The low stress required to move the boundaries is undoubtedly associated with the small shear, whether or not twinning dislocations are nucleated spontaneously. It should be noted that the calculation of the activation energy for nucleation using eqn. (31.17) neglects the core energy of the twinning dislocation, and this may not be justified when the elastic energy becomes small. The energy also does not include any term representing the misfit energy of the area enclosed by the loop of twinning dislocation; in other words, it is appropriate to the nucleation of a

new step on the surface of an existing twin, but not to the creation of a monolayer twin in a region of perfect lattice. However, if the calculation is correct, it seems that the normal growth of a twin which is only a few lattice spacings thick should present no difficulty, in contrast to the theory to be described below.

The conclusion that thickening of deformation twins in cubic crystals by spontaneous nucleation of successive loops of twinning dislocation is not possible rests on the assumption that the interface region possesses the elastic properties of the matrix. Several authors (see, e.g., Sumino, 1966; Delamotte and Alstetter, 1969) have pointed out that this conclusion may be erroneous and, in particular, the effective shear modulus,  $\mu'$ , to be used in eqns. (31.16) and (31.17) may be appreciably lower than the modulus,  $\mu$ , for bulk material. Some support for the view that the effective stress and activation energy required to nucleate a loop of twinning dislocation may be lower than the values given by the ordinary linear elastic model comes from the computer simulations of Yamaguchi and Vitek (1976) and Bristowe and Crocker (1977), which indicate that twinning dislocations have relatively wide cores and hence relatively small self energies. Yamaguchi and Vitek point out that the width of the core according to the Peierls-Nabarro model (pp. 273–276) is proportional to the shear modulus and, from the computed widths of twinning dislocations and lattice dislocations using various empirical interatomic potentials, they estimate the effective shear modulus  $\mu'$  to be a factor of three to five times smaller than the modulus of the matrix. Sumino estimated the stress needed for spontaneous twinning, i.e. virtual zero activation energy [see eqn. (31.16)], as  $\mu'b/2h$ , and Yamaguchi and Vitek point out that, with their potentials, this is  $\sim 0.01-0.02\mu$ . This is still rather a high stress for continuous growth of twins, but is lower than the corresponding estimate for formation of a twin nucleus from a screw dislocation, which is about  $0.03\mu$  for the same potentials. Thus growth by spontaneous nucleation of new layers is considered to be a possible mechanism, even in b.c.c. metals.

A theory for the normal growth of a twin without thermal activation was first given by Cottrell and Bilby (1951), and Millard and Thompson (1952) independently developed a similar description for the particular case of h.c.p. metals. The theory is analogous to Frank's theory of crystal growth from the vapour, and to the operation of a Frank-Read source, and is based on the topological properties of a node formed when a dislocation crosses from a parent into the twin (p. 292). This "pole dislocation" (e.g. *SNS* of Fig. 7.18) must have a Burgers vector with a component normal to the  $K_1$  planes equal in magnitude to  $qd$  or  $\frac{1}{2}qd$ . As the twinning dislocation (*BN* in Fig. 7.18) moves in the interface, it rotates about the node and (together with the node) is displaced through  $qd$  for each complete revolution. Successive blocks of structure,  $qd$  in thickness, thus suffer the same vector displacement relative to each other, to give the macroscopic shear of the twinning relation. Any necessary shuffling is not, of course, described by this mechanism, but presumably occurs spontaneously as the structure grows. Single nodes, giving a spiral step of twinning dislocation, and double nodes, giving closed terraces of twinning dislocation, are both possible, as in the associated problems of crystal growth and slip.

As discussed on p. 292, the geometrical conditions necessary for the setting up of a generating node of this type can always be fulfilled, and it is possible to envisage a number

of real or virtual processes by means of which single- or double-ended sources with the requisite configuration may be achieved. Such processes include compatible slip in a matrix–twin bicrystal, as in Fig. 7.18, glide of part of an expanding dislocation loop across an existing coherent matrix–twin interface so that the twinning dislocation ends on the two crossing points, partial transformation of a previously dislocated crystal, and the welding together of separate matrix and twin crystals along the  $K_1$  plane interfaces, through each of which an appropriate lattice dislocation emerges. A process which could be of particular physical interest is the expansion of a very thin disc-shaped twin or multilayer fault,  $nqd$  planes thick, so as to incorporate an existing lattice dislocation, or equivalently the glide of the dislocation from outside the periphery of the twin to a position where it intersects the twin disc. The limit of the twin will then be a continuous spiral of twinning dislocation running between generating nodes on the top and bottom surfaces, and rotation of the end elements in opposite directions about the pole dislocation will produce thickening of the twin upwards and downwards.

The generating node produced in this way requires the separate nucleation of at least a monolayer of twinning fault and, although the crystallographic and geometric conditions are relatively easy to satisfy, the physical operation of the source is subject to some restrictions. One of these arises from the fact that the twinning dislocations in the top and bottom surfaces have to pass each other at a separation of the fault thickness on the first revolution and this would require unrealistically high stresses if  $\mathbf{b}_T$  is appreciable and the fault is very thin. In particular, growth from a monolayer twinning fault is not likely to be feasible and estimates by Bilby (1953) indicate that, for cubic twins, a nucleus about 50 atoms thick would be required before the pole mechanism can begin to operate at experimentally observed twinning stresses. A second difficulty is that, in many lattices, the pole dislocation would not remain anchored, but would glide under the applied stress. The pole may be anchored firmly if it is a sessile partial rather than a lattice dislocation, but this is difficult to achieve. Nabarro (1952) showed that imperfect dislocations of Burgers vector  $q\mathbf{d}\mathbf{m}$  can exist only in special lattices and then are ineffective as they must lie in the composition plane. Dislocations with vectors  $2q\mathbf{d}\mathbf{m}$  can always exist, but are only useful if there is a second set of twinning planes at  $60^\circ$  to the first, as in b.c.c. crystals. In all other cases, the pole dislocations must be perfect dislocations in the two lattices. The stability of a node can then be due only to the difficulty which such a dislocation experiences in gliding; for example, dislocations with Burgers vectors of  $c[0001]$  are suitable as pole dislocations in some h.c.p. twinning modes, and probably exist in this structure although they do not contribute to glide processes.

Direct discussion of the very early stages of growth, i.e. of twin nucleation, is difficult, although several theories of growth have obvious implications in nucleation. In general, there are three types of nucleation theory, namely (1) twin nuclei are tiny stopped twins, pre-existing in the solid as a result of growth accidents (Oliver, 1952), (2) twins form spontaneously in the elastic field of an imposed external stress and/or in the field of internal defects, and (3) twins form by some kind of dissociation of existing dislocations or other defects, utilizing, for example, the large planar stacking faults which form readily in some f.c.c. alloys or the large displacements in the core of a b.c.c. dislocation which cannot

be treated by linear elastic theory. The experiments of Bell and Cahn and of Price apparently provide evidence in support of a combination of theories (1) and (2), but many later observations strongly suggest that twin nuclei in the common metallic structures form from dislocations.

The pole theories imply that the formation of an initial planar fault or monolayer twin will be more difficult than the addition of subsequent layers as the interfacial energy is not substantially changed after the first layer. Thus Cottrell and Bilby suggested that, if the stress resisting growth is  $X_i$ , that required for nucleation will be  $X_i + \sigma^f/b_T$ , where  $b_T$  is the magnitude of the Burgers vector of the twinning dislocation and  $\sigma^f$ , the fault energy, is approximately twice the energy  $\sigma$  of the  $K_1$  interface. If the fault forms by the bowing of a partial between two pinning points, an additional stress  $Cb_T\mu/y$  (where  $C$  is of the order of unity and  $y$  is the source length) is needed to overcome the line tension (see p. 268). The stress  $X_i$  incorporates any intrinsic resistance to motion (lattice friction or Peierls–Nabarro force) and also the effects of specific obstacles (other dislocations, faults, impurities, etc.) through which the twinning dislocation must pass. With an assumed source length of 25 nm, and a fault energy of  $40 \text{ mJ m}^{-2}$ , the required stress is only about 2.5 times the measured twinning stress in copper, so that a fairly modest stress concentration would appear to be adequate. Venables (1961) discussed in some detail the transition from load drops and unfaulted twins to smooth stress-strain curves with highly faulted twins which occurs in f.c.c. alloys as the stacking fault energy or temperature are reduced.

The simple conclusion that the nucleation stress exceeds the growth stress by only  $\sigma^f/b$  is in marked contrast to a calculation based on classical nucleation theory. In this theory, as in phase transformation theory, it is assumed that macroscopic concepts may be applied so that, for a sufficiently small twin, the surface energy of the twin matrix interface will be the dominant term in the net free energy change. When a very small twin forms, the free energy of the system should thus increase, and fluctuations may be necessary to overcome the lowest "saddle point" barrier. Homogeneous nucleation theory was applied to the formation of an enclosed twin embryo by Orowan (1954) and later re-examined by Christian (1969). The calculation is most acceptable if the twin embryos are assumed to be oblate spheroids parallel to  $K_1$ . The change in energy on forming an embryo of semi-axes  $R$ ,  $R$  and  $y$  then has a negative term representing the work  $\tau s$  per unit volume of embryo done by the shear stress across  $K_1$  resolved in the  $\eta_1$  direction, a positive term of magnitude  $Ay/R$  representing the elastic energy per unit volume of the constrained twin and a further positive term representing the (non-elastic) energy of the twin interfaces. In his calculation, Orowan treated the interface energy as that of the matching  $K_1$  planes of Fig. 20.17 plus that of the twinning dislocations assuming a uniform line tension; the constrained strain energy was treated as negligible on the assumption that the matrix stresses do not exceed the yield stress. An alternative approach is to allow this stress to approach the theoretical strength of the material on the basis that there will be no dislocation sources in the small volume around the tip of the nucleus. The strain energy must then be included, either by means of the Eshelby procedure or by treating the twinning dislocations as a pile-up, and the steps may be regarded as sections of a second interface with specific free energy  $\sigma'$  different from the

energy  $\sigma$  of the  $K_1$  interface. The change in energy  $\Delta G$  due to the formation of an embryo of volume  $v = 4\pi R^2 y/3$  is

$$\Delta G/v = -\tau s + 3\sigma/2y + 3\sigma'/R + Ay/R \quad (87.10)$$

The size and shape of a critical nucleus are given by  $\partial\Delta G/\partial R = \partial\Delta G/\partial y = 0$  and this defines the saddle point in  $\Delta G$ . The shape of this nucleus is defined by the equations

$$\tau s R = 2Ay + 3\sigma' \quad (87.11)$$

$$\sigma R = Ay^2 + \sigma'y \quad (87.12)$$

which may be solved to give the critical thickness

$$y_c = -(p + q) \pm (p^2 + q^2 - 4pq)^{1/2} \quad (87.13)$$

where  $p = -\sigma/\tau s$  and  $q = \sigma'/2A$ . [The minimization used by Orowan corresponds to  $A = 0$ , whereas some other treatments effectively have  $\sigma' = 0$ . The equations also depend slightly on the assumed geometry of the nucleus; that used by Orowan introduces factors of 4/3 and 3/4 into the second term on the right-hand side of eqn. (87.11) and the first term on the right-hand side of eqn. (87.12) respectively.]

The factor  $A$  is given in isotropic approximation by eqns. (52.26) and (52.27) and, for  $\{10\bar{1}2\}$  twinning in zinc, it is about  $6 \times 10^8 \text{ J m}^{-3}$ . Low estimates of  $\sigma$  and  $\sigma'$  are 20 and  $100 \text{ mJ m}^{-2}$  and, with Price's measured value of  $\tau = 500 \text{ MPa}$ ,  $\tau s = 7 \times 10^7 \text{ J m}^{-3}$ . The critical nucleus would thus have  $y_c \cong 0.63 \text{ nm}$  (i.e. about 10  $\{10\bar{1}2\}$  lattice planes) and the corresponding values of  $R_c$  and  $\Delta G_c$  are 15 nm and 75 eV respectively. This is clearly much too large an energy for homogeneous nucleation by thermal fluctuations to be feasible, and Price's original claim that the Orowan theory gives a nucleation barrier of only 1 eV seems to have been erroneous. Variations in the assumptions by setting either  $\sigma'$  or  $A$  equal to zero do not produce much improvement; the most favourable case is for  $A = 0$ , which corresponds to  $\Delta G_c \cong 13 \text{ eV}$  with the above assumptions, and an appreciably larger value with the effective value of  $\sigma' = 500 \text{ mJ m}^{-2}$  used by Orowan.

Homogeneous nucleation of twins is thus improbable unless there is a combination of very high stresses and very low surface and strain energies. Although the strain energy may have been reduced in the thin platelets used in Price's experiments, it is unlikely that the surface energies were substantially lower than the above estimates. It may be that the true stress concentration factor at the re-entrant nucleation sites was larger than that calculated or that the platelets contained undetected defects which aided nucleation. It is now considered probable that twinning in most metallic materials is initiated by other defects. Most nucleation models utilize the energy of one or more existing defects to make an initial single or multilayer fault which can then grow continuously or else link up with other similar faults to form a supercritical nucleus.

Continuous growth from a single-layer fault formed from the dissociation of an originally perfect dislocation in the parent lattice was originally proposed by Cottrell and Bilby (1951) for the b.c.c. structure and (independently) by Millard and Thompson for the h.c.p. structure. They suggested that a "generating node" (see p. 292) could form by the

dissociation of a single lattice dislocation, and thus avoid the need for separate nucleation. Unfortunately, there is a flaw in their analysis of this “pole mechanism” which may be best understood by considering first the general theory which was first advanced by Bilby (1953), and applied also to martensitic transformation. As already noted, Bilby and Christian (1956) considered a dislocation in the parent with Burgers vector  $\mathbf{b}_A$  which crosses the parent twin ( $K_1$ ) interface and continues in the twin as a dislocation with Burgers vector  $\mathbf{b}_B$  (see Fig. 7.18). They assumed that the two Burgers vectors are related by the twinning shear, i.e.  $\mathbf{b}_B = \mathbf{S}\mathbf{b}_A$ , and similarly that the slip plane normal  $\mathbf{m}'_B = \mathbf{m}'_A \mathbf{S}^{-1}$ . The components of the Burgers vectors, referred to unit cell-based coordinate systems in the twin and parent respectively are thus related by the correspondence matrix, as might be expected also if a fixed dislocation, without dissociating, is partly engulfed by a growing deformation twin. Note also that corresponding planes in parent and twin meet edge to edge in the  $K_1$  interface, which is a necessary geometrical condition for the propagation of slip across the interface. However, to avoid confusion, it should be noted at this point that some dissociations in which the Burgers vector of the dislocation within the twin has been given values other than  $\mathbf{S}\mathbf{b}_A$  have been postulated; in general,  $\mathbf{b}_B$  is then not a lattice vector of the twin.

The passage of the dislocation leaves a step in the interface of height  $h = \mathbf{b}_A \cdot \mathbf{m} = (b_A)_i m_i$  equal to the component of  $\mathbf{b}_A$  or equivalently of  $\mathbf{b}_B$  in the direction of the unit normal  $\mathbf{m}$  to  $K_1$ . This step is a twinning dislocation with Burgers vector  $\mathbf{b}_T = \mathbf{b}_B - \mathbf{b}_A = (\mathbf{S} - \mathbf{I})\mathbf{b}_A = (\mathbf{b}_A \cdot \mathbf{m})\mathbf{s}$  and it runs between two (opposite) crossing points or from a single crossing point to the periphery of the interface. Thus each crossing point is the junction of (at least) three dislocation lines, one in each crystal and one in the interface. Bilby called this configuration in which

$$\mathbf{b}_B = \mathbf{S}\mathbf{b}_A = \mathbf{b}_A + \mathbf{b}_T \quad (87.14)$$

a generating node. As all the planes parallel to  $K_1$  are now threaded by a “pole” dislocation with a normal component of Burgers vector, the set of parallel planes has been converted into a continuous spiral ramp leading from the parent phase to the twin. If the glissile twinning dislocation in the plane of the interface now glides along this plane whilst the other elements remain fixed, it must rotate about the node or nodes and, for each complete revolution, the  $K_1$  interface together with the node is displaced into the matrix or twin (depending on the sense of the rotation) through a distance  $h = \mathbf{b}_A \cdot \mathbf{m}$ . Successive blocks of the matrix structure thus suffer the same vector displacement relative to each other in order to give the macroscopic shear of the twinning relation and any necessary shuffles are accomplished spontaneously at the step in the interface. A dislocation which both enters and leaves a twin may do so at the same  $K_1$  interface or at the two  $K_1$  interfaces of a twin of finite thickness. Single nodes giving spiral steps from single crossing points and double nodes giving closed terraces of twinning dislocation from opposite crossing points in one interface can both be formed as in the topologically similar models of dislocation-assisted crystal growth and Frank–Read source operation. (Note especially the similarity to a cone source.)

The dislocation configuration required for this mechanism in compound or type I twins is quite specific; the pole dislocation must have a Burgers vector component normal to the interface which is equal to either  $qd$  (for  $q$  odd) or  $\frac{1}{2}qd$  (for  $q$  even) where  $d$  is the spacing of the lattice  $K_1$  planes. Any dislocation with a lattice Burgers vector from plane 0 to plane  $q$  or  $\frac{1}{2}q$  will satisfy this condition, which arises of course to ensure that  $\mathbf{b}_T$  when necessary shall be a zonal twinning dislocation and thus able to glide freely. The same condition ensures that if  $\mathbf{b}_A$  is a lattice vector of the parent crystal,  $\mathbf{b}_B$  is a lattice vector of the twin, and together with  $\mathbf{m}_B$  it defines a geometrically feasible, atomic slip system in the twin, the plane and direction of which may have relatively high indices, so that actual slip on the system may be impossible. Clearly, dislocation glide across the interface becomes a virtual rather than a physical process if the dislocation is unable to glide in the twin but, if the configuration can be formed in some other way, the resistance to slip could ensure the stability of the pole. However, in some lower symmetry crystals, there may not be any suitable pole dislocations.

It follows from this treatment that the interaction of independently formed stacking faults and dislocations may produce true generating nodes in the sense defined by Bilby and Christian (1956) and Sleswyk (1963) at each crossing point. In particular, four variables listed by Venables (1974) are correctly disposed so that the applied stress favouring twinning continues to drive the twinning dislocation into the matrix, whereas an opposite stress reverses the motion so as to produce detwinning. These variables are:

- (i) the screw sense of the pole dislocation;
- (ii) the type of stacking fault in the parent lattice which is trailed behind the glissile partial (or twinning) dislocation;
- (iii) the side of the fault on which the glissile partial is situated; and
- (iv) the direction in which the glissile partial moves under a fixed applied stress.

It may readily be seen that, if another pole dislocation has the opposite Burgers vector (i.e. the reversed screw sense), the last two variables must also be reversed and the left-handed rotation of  $-\mathbf{b}_T$  about the opposite pole will produce the same twinning shear as the right-handed rotation of the original twinning dislocation  $+\mathbf{b}_T$ .

The pole mechanism is usually invoked in combined nucleation and growth models of twinning which begin with the dissociation of a lattice dislocation. One simple possibility is that a glissile partial dislocation trailing a monolayer (twin type) stacking fault in the  $K_1$  plane encounters a suitable pole dislocation threading the plane. A pole source can then be produced if the partial wraps itself around the pole. The two parts of the partial then form a dipole on adjacent  $K_1$  planes, and between these two planes the pole dislocation is converted (formally) into the dislocation  $\mathbf{b}_B$  of eqn. (87.14). If the two opposite segments of the original partial are able to glide past each other (a very doubtful assumption with a single plane separation), they can continue to spiral upwards and downwards respectively, thus thickening the twin in both directions.

More elaborate descriptions suppose that an initial pole dislocation contains a superjog which dissociates in the  $K_1$  plane, emitting a twinning partial into this plane and leaving a

sessile partial along the original interface. With a correct configuration, this leads to a perfect pole in which the two ends of the twinning partial spiral upwards and downwards from the sessile partial, passing it very closely on the first revolution, but never being blocked by it. Thus a twin is formed by outwardly spreading shear on what is topologically a single helicoidal  $K_1$  interface wrapped around the pole dislocation. The sessile part of the dissociation is the vector  $\mathbf{b}_B$  referred to the parent coordinates and this becomes a lattice vector in the twin. Now suppose that the same pole dislocation dissociates into a dislocation with Burgers vector  $\mathbf{b}_B^*$  and a glissile partial  $-\mathbf{b}_T$  which glides away from the superjog in the opposite direction to produce an identical intrinsic, twin type fault. This dissociation is represented by

$$\mathbf{b}_B^* = S\mathbf{b}_A - 2(b_i m_i)\mathbf{l} = \mathbf{b}_A - \mathbf{b}_T \quad (87.15)$$

and variables (iii) and (iv) above have been reversed. The dislocation  $\mathbf{b}_T$  is still rotating about  $\mathbf{b}_A$  in the matrix but the vector  $\mathbf{b}_B^*$  is not a lattice vector because  $\mathbf{t}$  differs from the lattice vector  $S\mathbf{b}_A$  by  $2\mathbf{b}_T$ , which is not a lattice vector. Moreover,  $\mathbf{b}_B^*$  has the wrong screw sense as this has not changed whilst the vector of the twinning dislocation has been reversed, so that after one revolution the twinning dislocation does not advance along the pole but returns to the original sessile partial. If it recombines and then redissociates, it can continue only along its original path again, thus increasing the displacement between the same two planes and producing a high energy fault instead of a twin. This configuration, first pointed out by Cottrell and Bilby (1951) in connection with a model for f.c.c. twinning, was called an "anti-generating node" by Bilby. Twinning may in principle grow from such a node by utilizing intimate cross-slip in a manner first suggested by Venables (1961) for f.c.c. metals, and later described by Hirth and Lothe (1982) as a "ratchet" mechanism.

As mentioned above, there is a physical difficulty in the operation of a pure pole mechanism if the original twin is either a monolayer fault or has only a few layers. A generating node appears on both the top and bottom surfaces of the original fault, and the resulting twinning dislocation segments rotate in opposite senses. After one revolution, these two segments must pass one another on planes separated by the thickness of the original fault. For monolayer faults or thin twins, this would require a very large (static) stress. Bilby estimated that, for cubic twins, a nucleus about 50 atoms thick would be required before the pole mechanism began to operate at experimentally determined twinning stresses. Actually the separation of the partials is so small after the first turn that linear elastic theory may considerably overestimate the stress required; there is appreciable core overlap and atomistic calculations are needed. However, it does seem probable that the pole mechanism for growth may only be effective if a rather thick nucleus has developed in some other way, possibly by chance encounters of stacking faults with each other. In the analogous problem of the f.c.c. to h.c.p. transformation in cobalt, the situation is exactly similar except that the partials meet with a separation of two atomic layers. Seeger (1956) suggested that the partials may acquire sufficient kinetic energy to enable them to pass dynamically. This seems doubtful even for cobalt, where the effective

stress on the dislocations from the chemical driving force may be much larger than the externally imposed mechanical stress causing twinning.

Cottrell and Bilby's theory of b.c.c. twinning, although later shown to be incorrect, is instructive in clarifying the distinction between the two dissociations (87.14) and (87.15). They ingeniously utilized the fact that the twinning dislocation of type  $1/6\langle 111 \rangle$  is common to three crystallographically equivalent twin planes of type  $\{112\}$ . Figure 20.24 shows a length of dislocation, initially  $AOBC$ , lying in the  $K_1$  plane in which it cannot glide. Part of the line  $OB$  dissociates to form an initial stacking fault in the  $\{112\}$  plane, with nodes at  $B$  and  $O$ . This dissociation is

$$\frac{1}{2}[111] = \frac{1}{3}[112] + \frac{1}{6}[1\bar{1}\bar{1}] \quad (87.16)$$

Because  $BDEO$  is glissile, a length  $EO$  may lie along the  $[1\bar{1}\bar{1}]$  direction and hence is pure screw in nature. This length can then move into either the  $(\bar{1}21)$  planes or the  $(2\bar{1}1)$  planes which meet along the  $[1\bar{1}\bar{1}]$  direction. As it does so it generates a new fault on the plane into which it moves and Cottrell and Bilby originally claimed that it could rotate about  $OB$  whilst moving along  $OB$ , thus generating a twin along the stacking fault. As pointed out in the original paper, the dissociation (87.16) is energetically improbable and such faults are now believed to be mechanically unstable. The energy of the fault has to be supplied by the work done by the external stress [the elastic energy of the dislocations in eqn. (87.16) just balancing] and Cottrell and Bilby estimated the stress required as  $\sigma/b$ , where  $\sigma$  is the stacking energy. This stress is only  $\cong 0.001\mu$ , where  $\mu$  is the shear modulus, for materials of low fault energy but it has the rather large value of  $0.02\mu$  for  $\sigma = 200 \text{ J m}^{-2}$ , which might be an appropriate value for b.c.c. metals. Stresses of magnitude  $(10^{-3} - 10^{-2})\mu$  are often applied in order to twin b.c.c. metals so that, with a modest stress concentration, the theory seemed initially attractive. However, as first pointed out by Sleswyk (1974), the pole dislocation inside the twin does not have a lattice vector and there is a high energy fault inside the "twin". These defects were at first somewhat obscured by the initial decomposition and the utilization of the threefold symmetry around the  $\eta_1$  direction.

Hirth (1963) first distinguished pure pole mechanisms from what were later called ratchet mechanisms, and Hirth and Lothe pointed out a slight modification of the Cottrell-Bilby model which shows that, in fact, it is a ratchet. The initial configuration is similar to Fig. 20.24 but  $OE$  now represents an undissociated superjog in the dislocation so that there is no stacking fault on  $(112)$  and the lines  $CBOE$  and  $CBDE$  are coincident and form a section of perfect dislocation. Dissociation of the superjog  $OE$  takes place as according to eqn. (87.16) and the twinning partial  $OFE$  moves into either the  $(\bar{1}21)$  or the  $(2\bar{1}1)$  plane but, after a single turn, the parts of this loop which rotate about  $O$  and  $E$  respectively are both blocked by the sessile partial left along  $OE$ , so that the configuration cannot act directly as a pole source. This is illustrated in Fig. 20.25 which shows in (a) the stacking sequence  $\dots ABCDEF \dots$  in the matrix and the undissociated  $(1/2)[1\bar{1}\bar{1}]$  dislocation. The dissociation (b) is such that the partial dislocation leaves along 1 to the right of Fig. 20.25(b) and the two parts return along 2 and 3 to the left of the diagram. As they are opposites and very close together at this point, they may possibly annihilate

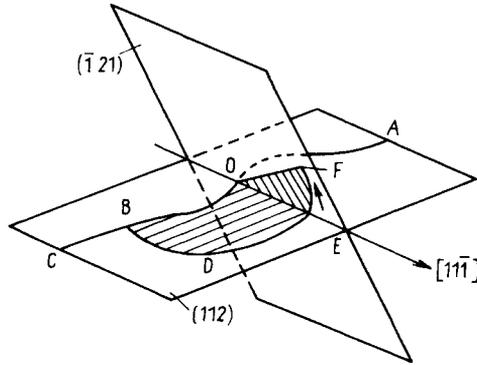


FIG. 20.24. The Cottrell Bilby "pole" mechanism for growth of b.c.c. (after Cottrell and Bilby, 1951).  $AO$  and  $OC$  represent lengths of lattice dislocation with a Burgers vector  $\frac{1}{3}[111]$ .  $OB$  is a sessile partial with a Burgers vector  $\frac{1}{3}[112]$  and  $BDEFO$  is a glissile twinning dislocation of Burgers vector  $\frac{1}{6}[11\bar{1}]$ . There is no dislocation along  $OE$ .

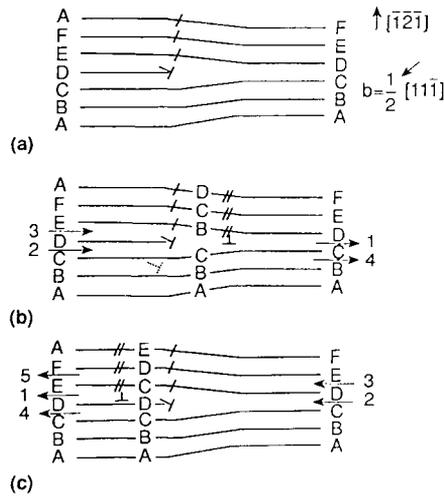


FIG. 20.25. The jog version of the Cottrell Bilby mechanism after Hirth and Lothe (1982) (a) the undissociated  $\frac{1}{3}[11\bar{1}]$  dislocation; (b) the configuration after dissociation into  $\frac{1}{3}[112]$  and  $\frac{1}{6}[11\bar{1}]$  dislocations; (c) the situation after the first revolution of the twinning dislocation.

over some length (formally by nucleation of a jog of unit height) leaving a closed loop to travel outwards and a residual part which is blocked by the sessile partial. Recombination with this partial will then produce the lattice Burgers vector and, if the reformed dislocation now cross-slips onto the next plane and then redissociates as before, a second loop of twinning dislocation is produced. Repeated operation of this cycle can thus (geometrically) produce a macroscopic twin. However, the mechanism requires the repeated dissociation and recombination of the lattice dislocation and this seems very improbable. Hirth and Lothe also pointed out that a dissociation into the opposite

twinning dislocation plus a larger sessile partial would give a true pole mechanism. The reaction is

$$\frac{1}{2}[111] = \frac{1}{3}[221] + \frac{1}{6}[\bar{1}\bar{1}1] \quad (87.17)$$

where  $1/3[221]$  is a lattice vector of type  $\langle 001 \rangle$  of the twin. Referring to Fig. 20.25(c), part of the twinning dislocation now leaves on path 1 and returns on path 3; it is not blocked but may leave again on 5 and continue to spiral upwards. The other part of the twinning dislocation leaves on path 1 and returns on path 2 and leaves again on 4 and continues to spiral in the opposite sense.

Cottrell and Bilby showed that their theory, if applied to a f.c.c. crystal, would produce only a monolayer of stacking fault, and at the time the paper was written this was in agreement with much experimental data in which twinning in f.c.c. materials was notably absent. Later, it was found that such twins are quite prevalent in alloys of low fault energy and Venables (1961) suggested a modified and ingenious mechanism for growth from a single stacking fault which was in effect the same ratchet mechanism described above for b.c.c. materials. Models involving single stacking faults seem much more reasonable for f.c.c. structures, where such faults actually occur, than for b.c.c. structures where they are never seen. Using the notation of the Thompson tetrahedron, consider the dislocation shown in Fig. 20.26 with a Burgers vector  $\mathbf{AC}$  lying in a plane except for a long jog  $N_1N_2$  lying in a plane  $a$ . Let the part of the dislocation in  $a$  dissociate into a Frank partial and a Shockley partial

$$\mathbf{AC} = \mathbf{A}\alpha + \alpha\mathbf{C} \quad (87.18)$$

Under the action of a stress the glissile Shockley partial  $\alpha\mathbf{C}$  moves away from the sessile Frank partial  $\mathbf{A}\alpha$  on the plane  $a$ , leaving an intrinsic fault [Fig. 20.26(b)] and, after reaching the unstable near-semicircular configuration, winds rapidly around  $N_1$  and  $N_2$  to reach the position shown in Fig. 20.26(c). Two segments of the  $\alpha\mathbf{C}$  dislocation delineating the fault meet along  $RS$  at a separation of only one interplanar distance, and a very large stress would be required to force these two opposite segments past each other. Note that the right-hand partial has moved downwards whilst wrapping around the pole dislocation so that it is the lower element of the dipole  $RS$ .

Venables assumed that the end element of the partial  $\alpha\mathbf{C}$  combines with the sessile partial  $\mathbf{A}\alpha$  along the length  $RN_2$  and the re-formed dislocation with Burgers vector  $\mathbf{AC}$  glides to the next plane and repeats the dissociation. When the second layer of fault expands, two opposite segments of dislocation will again appear along  $RS$ , but each will be displaced by one plane from the initial pair. One of these segments on the central plane will annihilate the twinning partial left there by the first dissociation, thus turning the two layers into one continuous fault wrapped twice around the pole dislocation, and the final configuration [Fig. 20.26(e)] is now a double helical layer of fault terminated by two partials  $SRN_1$  and  $SRN_2$  at a separation of two atomic layers. Repetition of these operations will lead to a twin limited by a helical twinning dislocation ending in segments parallel to  $RS$  on the top and bottom surfaces. Improbable though it seems, the ratchet geometrically works by converting the pole dislocation  $\mathbf{b}_B^*$  into the dislocation  $\mathbf{b}_B$  one plane at a time.

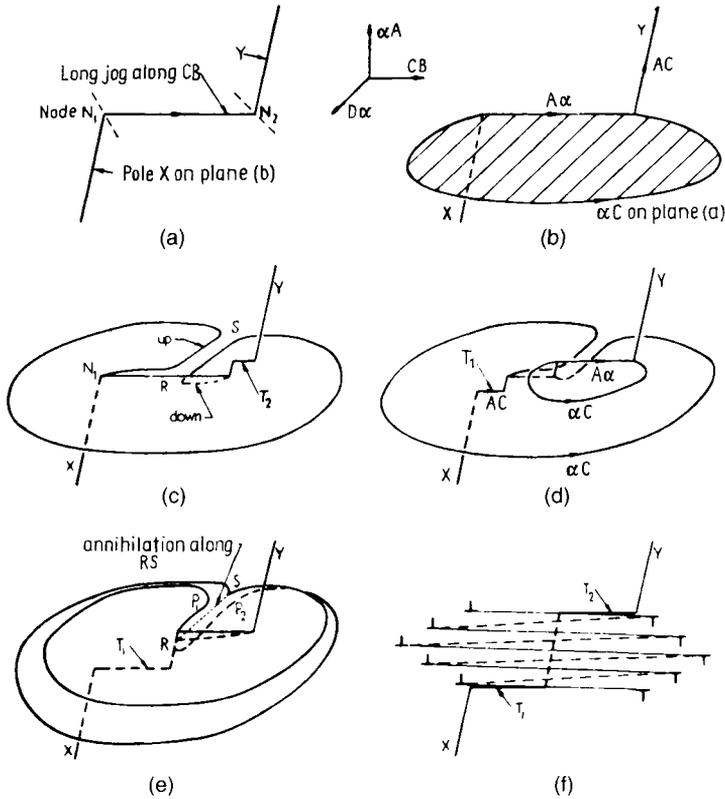


FIG. 20.26. Prismatic glide ("ratchet") mechanism for f.c.c. twin growth (after Venables, 1961). The notation is that of the Thompson tetrahedron. For details, see text.

The above discussion verifies Venables' claim, contrary to an objection by Sleeswyk (1974) that the pole dislocation in the twinned region is a perfect lattice dislocation. Venables also points out that, if only one generating node operates, a twin in the form of a planoconvex lens might be obtained. A rather similar situation is obtained when both crossing points of a given dislocation are on the same side (Fig. 20.27); once again the twin is planoconvex. Some limited experimental support for these latter models comes from the electron microscopic work of Steeds and Hazzledine (1964).

In the early discussion of Venables' theory, Hirth (1963) proposed an alternative dissociation to give a pure pole mechanism. However, Hirth's pole is not related to Venables' pole in the way that eqns. (87.14) and (87.15) are related, but involves instead dissociation of a different  $\frac{1}{2}(110)$  dislocation **AB** which has a Burgers vector at  $54^\circ$  to the antitwinning direction, whereas that of **AC** is at  $73^\circ$  to the twinning direction. Hirth's dissociation

$$\mathbf{AB} = \mathbf{A\alpha/BC} + \mathbf{C\alpha} \tag{87.19}$$

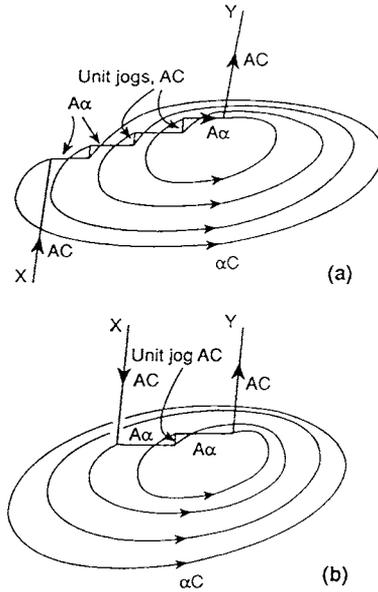


FIG. 20.27. Prismatic sources where the twinning dislocation remains in the matrix twin interface: (a) conventional prismatic source; (b) source from the end of a dipole which remains on one side of the twin to give a planoconvex shape (after Venables, 1974).

gives the opposite twinning dislocation  $C\alpha$  and a high energy sessile  $A\alpha/BC$  which is of type  $1/6\langle 411 \rangle$ . Another geometrically possible pure pole may be obtained by dissociating the lattice dislocation of eqn. (87.18) in the manner of eqn. (87.19) rather than that of eqn. (87.18). Suppose that the plane  $\alpha$  is a  $(111)$  plane and that the  $\alpha C$  direction is  $[11\bar{2}]$ . Equation (87.18) is then

$$\frac{1}{2}[110] = \frac{1}{3}[111] + \frac{1}{6}[11\bar{2}] \quad (87.18a)$$

and the corresponding pure pole is

$$\frac{1}{2}[110] = \frac{1}{3}[22\bar{1}] - \frac{1}{6}[11\bar{2}] \quad (87.20)$$

Similarly eqn. (87.19), which corresponds to

$$\frac{1}{2}[011] = \frac{1}{6}[141] - \frac{1}{6}[11\bar{2}] \quad (87.19a)$$

has a ratchet version

$$\frac{1}{2}[011] = \frac{1}{6}[\bar{1}25] + \frac{1}{6}[11\bar{2}] \quad (87.21)$$

The sessile products  $\mathbf{b}_B$  with Burgers vectors  $1/3[22\bar{1}]$  and  $1/6[141]$  have lattice Burgers vectors  $[001]_T$  and  $1/2[101]_T$  in the twin basis whereas the products  $\mathbf{b}_B^*$  with vectors  $1/3[111]$  and  $1/6[\bar{1}25]$  are non-repeat vectors  $1/3[111]_T$  and  $1/6[521]_T$  referred to the twin lattice.

All these dissociations are, however, energetically unfavourable and the elastic part of the energy varies in the order  $(87.20) > (87.21) > (87.19) > (87.18)$ . The dissociations will thus take place only if there are internal stress concentrations arising from pile-ups, intersecting slip or twin bands, or other agencies.

A notable feature of the pole and ratchet mechanism is that the parent dislocation which generates the twin has a Burgers vector not in the twin plane. This is described by Venables as a prismatic source. The special configuration required might be produced if AC is a slip dislocation which either has a straight length along CD or cross-slips into  $a$  and then lies along BC, or if jogs along BC are produced by dislocation intersections. Other theories which are based on prismatic sources include that of Fujita and Mori (1975), which utilizes the Cottrell–Bilby dissociation into Shockley and Frank partials but does not assume a pole mechanism. In contrast to the prismatic sources there are various theories in which the dislocation  $\alpha C$  is produced from dislocations BC or DC which lie in  $a$ , and these may be described as glide sources. Finally, there are theories which depend upon interactions of glide dislocations of two different systems, primary and coplanar (Mahajan and Chin, 1973a), or primary and cross-slip (Narita and Takamura, 1974). For clarity in subsequent discussion, the  $a$  plane will always be the twinning  $K_1$  plane, whereas the active slip plane will be some other plane for a prismatic source and  $a$  for a glide source.

Several investigations (Miura *et al.*, 1968; Narita and Takamura, 1974; Fujita and Mori, 1975) have shown that, for f.c.c. metals and alloys, twinning takes place on the most active slip plane (i.e. the primary or conjugate plane, depending on the amount of slip before twinning) and twins form only after appreciable slip. The twinning stress in a tensile test decreases as the orientation of the stress axis approaches  $\langle 111 \rangle$ , and also as the temperature is lowered, provided the twinning occurs in stage III of the work-hardening curve, but is little affected by orientation and temperature in stage II.

Early dislocation theories for the production of f.c.c. twins from glide sources usually depended on the formation of Lomer–Cottrell locks (Suzuki and Barrett, 1958; Haasen and King, 1960; Ookawa, 1957). In principle this is not necessary because an extended dislocation lying in its slip plane with its ends pinned may be separated into component partials by the action of an external stress. This will happen if the first partial is able to bow out under the applied stress, as in a Frank–Read source, whilst the back partial is unable to follow it because of its line tension. Venables has shown that the most readily separated dislocations of this type should favour twinning when the stress axis is near  $\langle 100 \rangle$ , and this is contrary to observation (see p. 912). The dissociation of other glide dislocations, which would give approximately the observed dependence of twinning tendency on orientation of the stress axis, is only possible with improbably large stresses and small source lengths.

The theories mentioned above attempt to overcome difficulties of this kind by utilizing Lomer–Cottrell barriers to anchor one of the partial dislocations. In the type of barrier discussed on p. 302, the faults are of intrinsic type and the apex of the barrier is away from the dislocation pile-up. Partial dislocations are thus unable to escape from the configuration. The theories imply that this situation is reversed, the apex pointing towards the pile-up, and the faults being extrinsic. The leading partial dislocations on each

plane can now glide away under the applied stress, but examination shows that this does not agree with the observed twinning directions; the “wrong” partials escape. It seems impossible to devise a source of this type which will produce the correct twinning partials unless a high energy stair-rod dislocation of type  $\alpha\mathbf{A}/\mathbf{B}\beta$  is included, and this should have little strength. More generally, Whelan (1958) has shown that the lock will have sufficient strength only if the Burgers vectors of the two interacting whole dislocations are at  $120^\circ$  to each other. This means that the resultant Burgers vector does not lie in the twin plane, so that the source as a whole is equivalent to the prismatic source discussed above.

Later theories are generally based on the experimental result that twinning in f.c.c. materials does not begin until slip is activated on at least two systems. The simplest description is that of Mahajan and Chin (1973a), who consider a reaction between dislocations of the primary system with Burgers vectors  $\mathbf{BC}$  and of the coplanar system with vectors  $\mathbf{DC}$  to form three Shockley partials

$$\mathbf{BC} + \mathbf{DC} = 3\alpha\mathbf{C} \quad (87.22)$$

which are then rearranged on successive planes to form a three-layer fault. A small twin is obtained when embryonic three-layer twins of this kind at different heights in a slip band grow together. Although the reaction (87.22) leads to an increase in energy (i.e. the undissociated dislocations repel each other), electron microscopic evidence on the production of extrinsic intrinsic fault pairs (Gallagher, 1966ab) suggests that it can occur by local attraction of the partials, without the stress concentration of a pile-up. More complex reactions involving slip activity on the cross-slip plane have been suggested by Narita and Takamura. Their model requires the reaction of a slip dislocation  $\mathbf{BC}$  with a Lomer dislocation  $\mathbf{DA}$  lying along the intersection of the primary and cross-slip planes to give

$$\mathbf{BC} + \mathbf{DA} = 2\alpha\mathbf{C} + \alpha\mathbf{A} \quad (87.23)$$

The Lomer dislocation  $\mathbf{DA}$  must be produced by reaction between dislocations of the cross-slip system with Burgers vector  $\mathbf{CA}$  and those of the coplanar system with Burgers vector  $\mathbf{DC}$ , and the theory has been criticized because such Lomer dislocations are seldom observed. However, the authors maintain that there is evidence of slip activity on the cross-slip plane, and that this means that the Lomer dislocations must form, even if not observed in lightly deformed material. This model was originally developed from a hypothesis that deformation twinning in f.c.c. materials may be viewed as a stress-relief process complementary to cross-slip, and it is claimed that it agrees well with the experimental results.

Yet another dislocation model (Fujita and Mori, 1975), described as stair-rod cross-slip, utilizes the same type of dissociation [eqns. (33.7) and (33.11)] as the original Cottrell-Bilby and Venables proposals, but simply assumes that a slip dislocation in (say) the primary system is held up by some obstacle along a direction  $\mathbf{DC}$  at  $60^\circ$  to its Burgers vector  $\mathbf{AC}$  and parallel to the intersection of the primary and conjugate slip planes, whereupon a Shockley partial  $\alpha\mathbf{C}$  cross-slips into the conjugate plane, leaving behind a Frank partial  $\mathbf{A}\alpha$  in place of the original lattice dislocation. This gives a wide stacking fault

on the conjugate plane, but this fault is not considered to wrap around the remaining parts of the original dislocation (as in the Venables model) but simply to extend away from this dislocation. Successive slip dislocations are then assumed to pile up behind the Frank partial and to cross-slip in the same way, thus forming an array of stacking faults on successive conjugate planes, i.e. a thin twin. Thus if the tensile axis lies in the standard unit triangle  $[001]-[011]-[111]$  the primary Burgers vector in the  $(111)$  slip plane is  $a/2[\bar{1}01]$  and the dissociation is

$$a/2[\bar{1}01] = a/6[\bar{1}21] + a/3[\bar{1}\bar{1}1] \quad (87.24)$$

where the partial  $a/6[\bar{1}21]$  glides away on the conjugate  $(11\bar{1})$  plane. However, if the stacking fault energy is reasonably low, the original slip dislocation will be dissociated into Shockley partials  $a/6[\bar{2}11]$  and  $a/6[\bar{1}\bar{1}2]$ , and the partial which cross-slips onto the conjugate plane will form by the further dissociation of the leading partial

$$a/6[\bar{2}11] = a/6[\bar{1}21] + a/6[\bar{1}\bar{1}0] \quad (87.25)$$

so that a low energy stair-rod is left at the junction of the two slip planes.

As previously discussed in relation to the Cottrell–Bilby theory, there is an increase in energy associated with the reactions (87.24) and (87.25) so that it is not obvious that stair-rod cross-slip will occur in practice; nevertheless, the simplicity of this proposal, which is in sharp contrast to the complexity of the Venables model, is appealing. It can be equally applied to cross-slip from the conjugate plane to give a microtwin on the primary plane, and it appears to be consistent with many of the experimental observations on f.c.c. twinning. In their original paper, Fujita and Mori (1975) pointed out that, because of conjugate or primary slip prior to stair-rod cross-slip, the twinning plane may already contain many Shockley partials with the same Burgers vector as the cross-slipped partial. If these interact strongly with the cross-slipped partials, it may be difficult for the latter to glide over large distances to create the successive faults which constitute a twin. They therefore suggested an alternative cross-slip process between coplanar slip bands on the two planes, leading to twinning partials which can easily glide. However, as found by other workers (see p. 912) and confirmed by Mori and Fujita (1977), the observed twinning direction always corresponds to the Burgers vector of the leading Shockley partial of the conjugate or primary slip direction, as predicted by the simple theory without coplanar slip.

Mori and Fujita (1980) were subsequently able to make direct electron microscopic observations (including Burgers vector analyses) of the stair-rod cross-slip of eqn. (87.25) in foils taken from single crystals of a Cu–11 at.% Al alloy deformed in tension at 77 K. The composition and test temperature were chosen to give a slow twin growth rate so that the wide overlapping stacking faults which are considered to be the initial stage of twinning could be directly observed. Observations of the progressive accumulation of stacking faults by stair-rod cross-slip on successive planes have also been made on Cu–8 at.% Ge single crystals during in-situ straining in a 3 MV electron microscope (Mori *et al.*, 1981). As the twin formed, the stair-rods and the trailing Shockley partials (or the Frank partials which represent their combined effect) accumulated at the intersections between

the primary slip bands and the twin interfaces. Other primary slip bands were found to act as obstacles to the extension of thin twin plates along the conjugate plane but, once such a band has been penetrated by the twin, its dislocations are able to contribute to thickening of the twin by the above mechanism. These results apparently confirm the Mori–Fujita growth mechanism in thin foils, but its importance in the twinning of bulk specimens remains to be established.

Now consider dislocation mechanisms which have been suggested for twin formation in b.c.c. metals and alloys. An obvious difference is that wide single-layer stacking faults have not been observed to form in any b.c.c. metal or alloy, and indeed are thought to be mechanically unstable (Vitek, 1968). Most theories of twin nucleation in this structure thus depend essentially on rearrangements within the core of a dislocation to give a multilayer twin. In particular, Sleswyk (1963) proposed that, because of the threefold symmetry of an unstressed  $a/2(111)$  screw dislocation, it may be regarded as having a three-dimensional core with a  $a/6(111)$  partial on each of the intersecting  $\{112\}$  planes. Under stress, however, this configuration will be unstable, and the partials could rearrange to form a three-layer twin on the most highly stressed of the  $\{112\}$  planes. A very similar suggestion was made by Ogawa (1965), who considered that edge dislocations might spread the total Burgers vector onto three successive planes, and thus give a three-layer fault. Such speculations appear to be consistent with the early  $\gamma$ -surface calculations (Vitek, 1968) which indicated that very thin “twins” [strictly four-layer faults as the interfaces were of the “isosceles” type (see p. 919)] may be mechanically stable.

Interest in the core structure of the b.c.c. screw dislocation has been greatly stimulated by the important role assigned to such dislocations in the low temperature deformation behaviour (Christian, 1970, 1983; Vitek, 1985), and many computer simulations of the structure and its change when subjected to shear and/or dilatational stresses have been made. These calculations show that the core structure is quite complex and is sensitive to the (two-body) interatomic potential and to both shear and non-shear components of the stress tensor. In certain circumstances, the application of stress leads to the formation of a four-layer fault, i.e. an apparent twin embryo, rather than to deformation.

The main deformation twins in h.c.p. materials (see Table XVIII) have  $K_1$  planes of type  $\{10\bar{1}2\}$  (all metals) and  $\{11\bar{2}1\}$  and  $\{11\bar{2}2\}$  (titanium group metals). In addition,  $\{11\bar{2}4\}$  twins which are conjugate to  $\{11\bar{2}2\}$  twins have been found in magnesium and titanium and there are several other modes, notably the conjugate  $q=8$  modes  $\{10\bar{1}1\}$  and  $\{10\bar{1}3\}$  observed in titanium and magnesium respectively. Transformation twinning of either type I with  $K_1 = \{10\bar{1}1\}$ ,  $q=4$ , or its type II equivalent is frequently observed after the b.c.c.–h.c.p. martensitic transformation.

Yoo (1981) has reviewed twinning in h.c.p. materials and Fig. 20.28 shows his plot of twinning shear versus  $\gamma = c/a$  ratio for various twinning modes, with the observed modes of seven h.c.p. metals superimposed. The  $\{10\bar{1}2\}$  mode is found in all cases, despite the shear reversal at  $\gamma = 3^{1,2}$  already noted.

For a uniaxial stress applied along the  $c$  axis, twins of a particular mode may form if the mode line in Fig. 20.28 has a negative slope, whilst a compressive stress will be able to

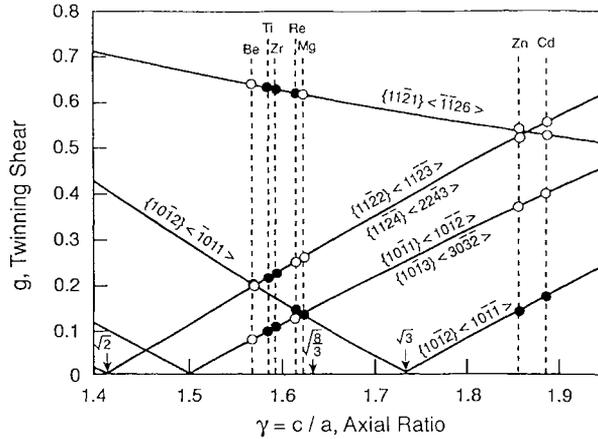


FIG. 20.28. Variation of the twinning shear with the axial ratio of the h.c.p. structure (after Yoo, 1981). The seven h.c.p. metals are indicated in the Figure by filled and open circles; a filled symbol indicates that the mode is active.

utilize only those modes on a line with a positive slope. This rule is reversed for the two conjugate modes listed on the same plot as their primary modes. Thus, with respect to the *c* axis, the  $\{11\bar{2}1\}$ ,  $\{11\bar{2}4\}$  and  $\{10\bar{1}3\}$  twins are tension twins and the  $\{11\bar{2}2\}$  and  $\{10\bar{1}1\}$  twins are compression twins. The  $\{10\bar{1}2\}$  twin is a compression twin for zinc and cadmium and a tension twin for all the other metals.

Millard and Thompson (1952) independently suggested a pole mechanism for the growth of  $\{10\bar{1}2\}$  twins. As this is a  $q=4$  mode, the expected twinning dislocation is a zonal dislocation of double step height with a Burgers vector given by eqn. (87.6), where the factor  $(3 - \gamma^2)/(3 + \gamma^2)$  is  $+(1/17)$  for ideal  $c/a$  and  $-(1/13)$  for zinc which has  $c/a \cong (7/2)^{1/2}$ . Thompson and Millard did not explicitly consider growth from a single stacking fault, but treated the case where a separate fault and dislocation intersect; in any event, the node formed should be written

$$\mathbf{b}_B = [10\bar{1}0]_T = [0001]_P - (3 - \gamma^2)/(3 + \gamma^2)[10\bar{1}1] \tag{87.26}$$

Evidence for twinning dislocations in a  $\{10\bar{1}2\}$  interface in zinc was first obtained by Antonopoulos *et al.* (1989) but the authors estimated the Burgers vector as about  $\frac{1}{4}[10\bar{1}1]$ , which corresponds to a step of about six to eight lattice planes instead of the expected two. There is no obvious reason for this and, in other work (Lay and Nouet, 1994; Braisaz *et al.*, 1996), the correct twinning dislocations have been identified by high resolution electron microscopy. In addition, various other linear defects in the interface were identified. The defect character of these lines was determined by Pond's circuit operator and by mapping closed circuits around the defect into a reference lattice. In addition to the twinning dislocations discussed above, other dislocations with large steps (of the order of  $15-17d$ , where  $d$  is the spacing of the  $\{10\bar{1}2\}$  planes) but with small Burgers vectors were also identified; they were all consistent with the predictions of Pond's theory (see p. 923).

Bacon and Serra (1994) have proposed a new mechanism for the thickening of a  $\{10\bar{1}2\}$  twin as a result of their computer simulation. It arises from the interaction of a lattice basal plane dislocation with its  $1/3\langle 11\bar{2}0 \rangle$  Burgers vector at  $60^\circ$  to the interface. The dislocation does not glide into the twin, propagating slip across the interface, but instead decomposes into interfacial defects. Specifically it forms  $m$  twinning dislocations plus a residual interface dislocation and the twinning dislocations can then glide in the interface, thickening or thinning the twin. Actually, the usual value of  $m$  is three, so that rather a large number of matrix dislocations would have to interact in order to get a twin of reasonable macroscopic thickness, and this is only a growth law; nevertheless, it is a mechanism which obviates the need for a pole mechanism and is thus of considerable interest.

The Burgers vector of an elementary twinning dislocation for the  $\{11\bar{2}1\}$  mode in cobalt is about  $(1/35)\langle 11\bar{2}\bar{6} \rangle$  and Vaidya and Mahajan (1980) suggested that the following reaction of two  $\frac{1}{3}\langle 2\bar{1}1\bar{3} \rangle$  dislocations with a  $\langle 1\bar{1}00 \rangle$  dislocation would yield a multilayer stacking fault approximating to a thin twin:

$$2(1/3)\langle 2\bar{1}1\bar{3} \rangle + \langle 1\bar{1}00 \rangle = 12(1/36)\langle \bar{1}\bar{1}2\bar{6} \rangle$$

It was suggested that the  $\langle 1\bar{1}00 \rangle$  dislocation might arise from interactions between dislocations with  $\mathbf{a}$  type Burgers vectors

$$(1/3)[2\bar{1}\bar{1}0] + (1/3)[1\bar{2}10] = [1\bar{1}00]$$

This mechanism is similar in concept to that suggested by Mahajan and Chin for f.c.c. twinning, although the  $1/3\langle 2\bar{1}1\bar{3} \rangle$  dislocations and the spontaneous spreading of the product into 12 adjacent planes do not seem very probable. Moreover, in the case of  $\{11\bar{2}1\}$  twinning it was noted above that the elementary twinning dislocations are probably split into steps of atomic height which are thus partial twinning dislocations (supplementary displacement dislocations) with Burgers vectors of about  $(1/70)\langle 11\bar{2}\bar{6} \rangle$  (see p. 926).

A true pole mechanism can be envisaged in principle for  $\{11\bar{2}2\}$  twinning, the pole dislocation having a Burgers vector of type  $\frac{1}{3}\langle 1\bar{2}1\bar{3} \rangle$ . Clearly, this appears improbable, as do specific proposals for other h.c.p. twins.

Table XX shows some results for the computed energies of the relaxed  $K_1$  interfaces and for various structural and energetic features of steps (i.e. twinning dislocations) in these interfaces. The last column shows the applied shear strain at which the twinning dislocation was displaced along the interface. Some twinning dislocations were found to have narrow (three-dimensional) cores  $\sim 1-2a$  in diameter, and these did not move until the applied shear strain reached  $\sim 1-4\%$ . Others have wide (planar) cores  $4-6a$  in diameter and glide along the  $K_1$  interface at much smaller applied strains. The highly glissile steps are those which correspond to the  $\{10\bar{1}2\}$  and  $\{11\bar{2}1\}$  twins, and the steps of low mobility are those on the  $\{11\bar{2}2\}$  and  $\{10\bar{1}1\}$  interfaces. The difference in mobility of the twinning dislocations in  $\{10\bar{1}2\}$  and  $\{10\bar{1}1\}$  is striking in view of the apparent similarity of their interface structures; it is believed to be related to the magnitude and complexity of the atomic shuffles. In fact, the mobilities of the steps of height  $d$  in the  $\{11\bar{2}1\}$  interface and  $2d$  in the  $\{10\bar{1}1\}$  interface, corresponding to the unobserved  $\{11\bar{2}1\}$  shear mode and the

TABLE XX. COMPUTED ENERGIES AND PROPERTIES OF SOME h.c.p. TWIN INTERFACES (after SERRA *et al.*, 1991)

$K_1$	$\gamma$ ( $\varepsilon/a^2$ )	$b_T$ (eq.)	$s$	$b^2/a^2$	$h/d$	Disl. line	Energy elast.	( $\varepsilon/a$ ) core	Core width	Critical strain
10 $\bar{1}2$	1.15	38a	0.12	1/51	2	0.1	0.1	0.0	6a	0.002
11 $\bar{2}2$	0.92	39a	0.27	4/33	3	3.7	1.6	2.1	a	0.014
11 $\bar{2}1$	0.73	40a	0.61	3/140	$\frac{1}{2}$	0.3	0.3	0.0	11a	0.001
10 $\bar{1}1$	0.64	41a	0.15	25/123	4	4.0	2.3	1.7	a	0.02
		42	0.36	37/123	2	3.6	2.4	1.2	2a	0.006

Notes: This table is adapted, with minor corrections, from Tables 1 and 2 of Serra *et al.* 1991. Only observed deformation modes plus the observed  $\{10\bar{1}1\}$  transformation twinning mode are listed. The values of  $s$  and of  $b^2/a^2$  apply to the ideal axial ratio,  $\gamma = (8/3)^2$ . The energy unit  $\varepsilon$  is the depth of the atomic pair potential at the nearest neighbour distance.

transformation mode previously discussed, were found to be respectively very much higher and slightly higher than those of the steps in the observed modes.

Semiconductors such as silicon and germanium have relatively immobile dislocations because of high Peierls–Nabarro forces. At low temperatures, brittle fracture occurs under stress with little evidence of plastic deformation but, if deformation is obtained below the brittle–ductile transition temperature of  $\sim 550$  C, for example by superimposing a uniform hydrostatic pressure on the specimen, deformation takes place largely by twinning. The diamond structure, as noted earlier, consists of two interpenetrating f.c.c. lattices displaced relative to each other by  $(1/4)a(111)$  so, in terms of the stacking of close-packed layers, the structure is...  $A\alpha B\beta C\gamma$ ... A distinction has to be made between slip on a plane midway between the closely spaced  $\alpha$  and  $B$  layers and slip on a plane midway between the widely spaced  $A$  and  $\alpha$  layers. The two sets of dislocations are called the glide and shuffle sets respectively.

A mechanism for twinning on the  $\{111\}$  planes has been suggested by Pirouz (1987, 1989) and by Pirouz and Hazzledine (1991). It is an important feature of the model that, for both screw and 60° dislocations, the dissociation into partials leaves the leading partial much more mobile than the trailing partial. Consider the dissociation of part of a screw dislocation of the glide set with a Burgers vector  $\mathbf{BA}$  dissociated into partials  $\delta\mathbf{A} + \mathbf{B}\delta$  on plane  $a$  between pinning points  $H$  and  $H'$  (see Fig. 20.29). The leading partial of high mobility may escape from the trailing partial and, by winding around  $H$  and  $H'$ , may eventually approach  $\mathbf{B}\delta$  again from the other side. Pirouz originally assumed that, despite their mutual repulsion, the two partials could be forced together again to re-form  $\mathbf{BA}$ , which could then cross-slip onto the next plane and repeat the process. Clearly this is a ratchet mechanism very similar to those already discussed. This exact mechanism was considered for f.c.c. twinning but was rejected because it was thought that the stress needed to separate the two partials would be too high. The model clearly works better for the diamond and zinc blende structures where quite large differences in the mobilities of the partial dislocations have actually been observed, but it suffers from the inherent improbability of all ratchet mechanisms and the difficulty of fast growth. The

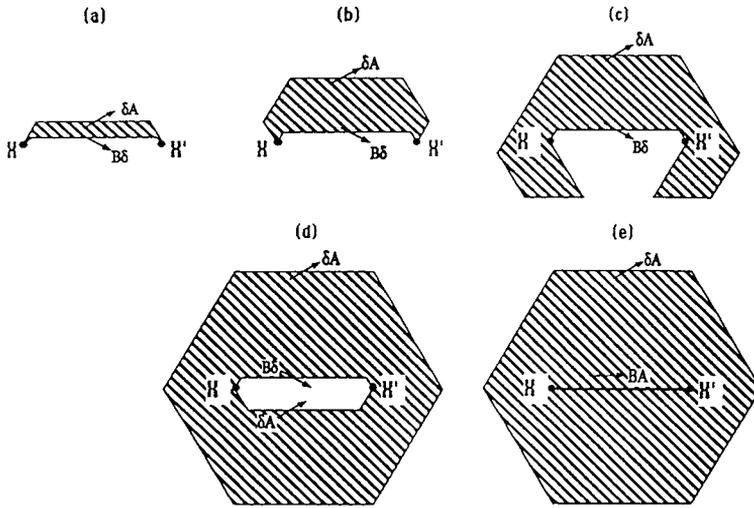


FIG. 20.29. Schematic diagram of the model proposed by Pirouz and Hazzledine (1991) for the formation of (111) twins in diamond cubic and zinc blende structures.

recombination difficulty can be avoided, at least in part, by adopting a cross-slip mechanism due to Friedel and Escaig which allows the extended configuration to cross-slip without recombining over its whole length and which can operate without any resolved shear stress on the cross-slip plane.

### 88. EFFECTS OF EXTERNAL VARIABLES ON TWINNING

This section examines very briefly the factors affecting twinning; for more detail, see Christian and Mahajan (1995). Many of the effects are poorly understood at present and, in such cases, the experimental results will be presented without comment.

The question of a critical resolved shear stress for twinning was discussed (inconclusively) above; the data are not accurate enough or sufficiently reproducible to decide. What is always observed is that the first twin to form is on the highest stressed plane of the particular twinning mode, thus confirming that the shear stress is important.

The effect of temperature is to increase the tendency to form twins as the temperature is reduced. In a few cases, twins will only form at liquid helium temperatures. Early, very careful work by Suzuki and Barrett (1958) on single crystals of silver-gold alloys of varying composition but fixed orientation established three regimes (see Fig. 20.30); in the high temperature region I, a localized band of twins is formed on the primary or conjugate slip planes and spreads across the specimen into two opposite quadrants. This is followed by a second band of twins which spreads into the other two quadrants. Twinning is accompanied by load drops. In region II, twin bands form on either the primary or the conjugate planes, depicting stress fields which are opposite to those given by the twin

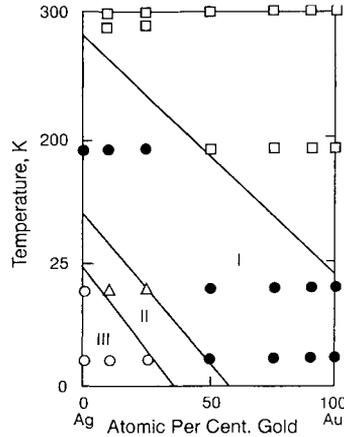


FIG. 20.30. Temperature-concentration diagram showing the occurrence of twinning in silver-gold alloys. In domain I, twinning occurs on the primary slip plane. In domain II, twins form on both primary and conjugate planes, but in different regions of the specimen. In domain III, the two types of twin coexist on the same planes.

bands. The twins grow until they impinge on one another. In region III, found only in silver-rich alloys at low temperatures, twins form copiously on both primary and conjugate planes.

Twin formation is also encouraged by low temperatures in b.c.c. structures, but often occurs here at the beginning of the deformation with large load drops. It is not possible, in general, to measure the stress at which twinning begins, just as it is impossible to measure a true upper yield stress.

Temperature effects are usually correlated with strain rate effects through some kind of chemical rate theory equation. But not here. The fracture toughness seems little sensitive to the strain rate at temperatures where the temperature variation is steepest. However, the strain rate can have dramatic consequences: normally ductile metals (even aluminium foil) are often brittle when shock loaded, but are perfectly ductile at more conventional strain rates.

The lower yield stress for twinning often follows a Hall-Petch relation

$$\sigma_y^T = \sigma_0 + k'd^{-1/2}$$

where  $d$  is the grain diameter,  $\sigma_y^T$  is the yield stress at which twinning begins and  $\sigma_0$  and  $k'$  are material constants. A conventional explanation of the usual Hall-Petch equation is based on the assumption that pile-ups accumulate against a grain boundary and, at yielding or flow, the stress field at the head of the pile-up must exceed some critical value which allows a twin to form.

Not all results support the Hall-Petch equation. Vohringer (1970), for example, found a good Hall-Petch relation for Cu-5 at.% Sn alloy but, for a Cu-15 at.% Zn alloy over a similar range, the twinning stress was linear in  $d^{-1}$  rather than  $d^{-1/2}$ .

Twinning as a deformation mechanism is diminished and ultimately removed by interstitial solutes in b.c.c. materials. According to Magee *et al.* (1971), this may be a simple consequence of the crystallography of twinning. The disordered b.c.c. solution has interstitial atoms at random on the octahedral sites  $\frac{1}{2}\langle 100 \rangle$  and  $\frac{1}{2}\langle 110 \rangle$ , and only one-third of these sites become octahedral sites of the product structure. The shear translates the other interstitial atoms to sites like  $\langle 00\frac{1}{2} \rangle$  and  $\langle \frac{1}{4}\frac{1}{4}\frac{1}{4} \rangle$ , and this gives a formidably large energy which must be reduced by shuffling.

Substitutional solids generally increase the tendency to twin, although the effects vary from one solute to another. Bolling and Richman (1967) found that twinning was the main deformation mode in iron–beryllium alloys with about 25% beryllium, and many iron alloys showed a transition from slip to twinning when the solute concentration exceeded a certain value. Twinning was the initial mode of deformation at 77 K at a slightly lower stress than at room temperature. The stress versus strain curves, without load drops, were ascribed to “continual mechanical twinning”.

Several solutes apparently promote twinning in the transition metals; the outstanding example is rhenium. Substitutional solutes also have large effects on the copper group metals, largely because of the big reduction in stacking fault energy which accompanies solution in these alloys. Figure 20.31 is an old plot by Venables of twinning stress versus stacking fault energy. As cobalt alloys with higher stacking fault energies twin even more readily than the copper alloys, it has been suggested that, in f.c.c. alloys, twinning is favoured not by low fault energy but by a tendency to planar slip, as found by Gerold and Karnthaler (1989) and by Hong and Laird (1990).

A number of observations show that twinning may be suppressed by a strain previously supplied at a higher temperature. Twinning on shock loading can be prevented in this way; the amount of prestrain required increases with decreasing temperature of the test and also depends on the final strain rate. The experiments of Boucher and Christian (1972) also indicated that a homogeneous distribution of screw dislocations is more effective than a heterogenous structure.

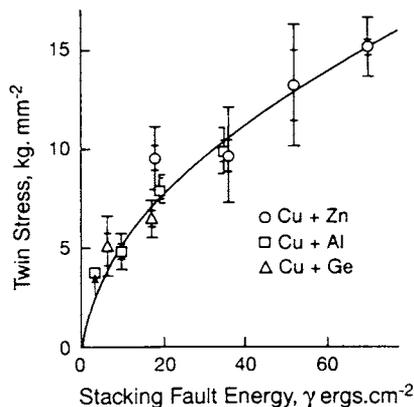


FIG. 20.31. The dependence of twinning stress on stacking fault energy for copper-based alloys (after Venables, 1974).

Precipitate particles and small inclusions generally have an inhibiting effect on twinning. In the case of partly coherent precipitates, the probable reason for this is the difficulty of forcing the twinning shear through the particle.

Some problems associated with the compatibility of slip and twinning will be examined briefly. For both cubic structures, parallel slip bands and twin lamellae are clearly geometrically possible as the  $K_1$  plane is parallel to the usual slip plane in f.c.c. and to a frequently observed (albeit not the most favoured) slip plane in b.c.c. As noted above, this also leads to the accommodation of internally terminating twin lamellae by emissary slip, in the simplest form of which the slip is homogeneous over a band corresponding to the thickness of the twin. Mahajan (1971) has shown that in simple examples of emissary slip in molybdenum–rhenium alloys, the Burgers vectors of slip and twinning dislocations are indeed parallel.

Now consider the propagation of slip across a coherent  $K_1$  interface of a twin. The simplest case is when the interface contains the Burgers vector of the slip dislocations, as dislocations on the intersecting slip plane must then meet the interface in screw orientation. They can thus cross-slip into the twin on any convenient slip plane of the latter without leaving a defect in the matrix–twin interface. When this condition is not satisfied, the matrix slip can still often be continued into the twin on any plane which meets the original slip plane edge to edge in the interface, but conservation of the Burgers vector then means that a linear defect must be left in the interface. If the Burgers vectors of the slip dislocations in parent and twin are corresponding vectors,  $\mathbf{b}_A$  and  $\mathbf{b}_B$ , so that they are related by

$$\mathbf{b}_B = \mathbf{S}\mathbf{b}_A = \mathbf{b}_A + \mathbf{b}_I \quad (88.1)$$

where  $\mathbf{S}$  is the twinning shear and  $\mathbf{b}_I$  the Burgers vector of the residual interface defect, their two slip planes are necessarily corresponding planes which automatically meet edge to edge. This is the case already considered in connection with the pole mechanism; if a finite length of the matrix dislocation crosses the interface, the linear defect joining the two crossing points is a twinning dislocation. Note that, if the slip is to propagate into the twin, the new Burgers vector must be a repeat vector of the twin lattice, and this requires that the step height in type I twinning must be  $qd$  or  $\frac{1}{2}qd$ ; i.e. if  $q > 2$ , the defect left in the interface must be a zonal twinning dislocation (Saxl, 1968). This is the same condition as that given on p. 937 for the formation of a generating node; there is an equivalent condition for slip propagation into a type II twin.

A lattice dislocation of the parent meeting  $K_1$  in a non-screw orientation will always have a corresponding twin vector which is either a lattice vector or an integral fraction of a lattice vector. In the latter case, the dislocation will be unable to cross the interface without leaving a stacking fault in the twin, but it will always be geometrically possible for a group of such dislocations to cross simultaneously, forming a single lattice dislocation in the twin. An example of this which has been studied experimentally is slip across a  $\{10\bar{1}2\}$  twin in a h.c.p. structure. For a given twin plane, only one of the three  $\langle 11\bar{2}0 \rangle$  slip vectors of the parent structure is transformed by the twinning shear into a lattice vector of the twin, and there is experimental evidence that a moving twin boundary pushes dislocations of the other two

types ahead of itself, rather than incorporating them as faults. As this twin mode has  $q=4$ , pairs of dislocations of the parent lattice always transform into whole dislocations of the twin; in the case of the two  $\langle 11\bar{2}0 \rangle$  close-packed vectors mentioned above, the corresponding twin vector is of type  $1/3\langle 11\bar{2}3 \rangle$ . The slip plane of the parent may be either the basal plane (0001) or a prismatic plane of type  $\{1\bar{1}00\}$ , and the corresponding slip planes of the twin are respectively of types  $\{1\bar{1}00\}$  and  $\{11\bar{2}2\}$  (Yoo and Wei, 1966). Tomsett and Bevis (1969) were actually able to observe basal plane slip in thin foils of zinc crossing a  $\{1\bar{1}02\}$  twin boundary and continuing on the  $\{1\bar{1}00\}$  plane of the twin, despite the unusual Burgers vector of the twin slip system (see Fig. 20.32).

Dislocations crossing a twin interface, even in non-screw orientation, are not geometrically constrained to continue on the corresponding slip plane; any plane which contains the line of intersection of the parent slip plane and the  $K_1$  interface is a possible new slip plane and the Burgers vector of the new slip system is obtained by adding an appropriate twin lattice vector to  $\mathbf{b}_B$  to give a net vector in the new slip plane. Clearly the same addition has to be made to  $\mathbf{b}_T$  in eqn. (88.1), so that the new plane is obtained at the expense of a more complex interface dislocation and (probably) of a large slip vector. Finally, it should be emphasized that geometrical compatibility of slip across a twin interface does not imply that it will necessarily occur. The new slip system may not be a usual slip system of the crystal structure, or it may have too low a resolved shear stress acting on it.

The propagation of a deformation twin (*A*) across an existing twin (*B*) forming a secondary twin (*C*) in the crossed region may be examined by considering the compatibility of the various twinning shears, or by extending the above theory of crossing slip to the case where the crossing dislocation has a partial (i.e. twinning) Burgers

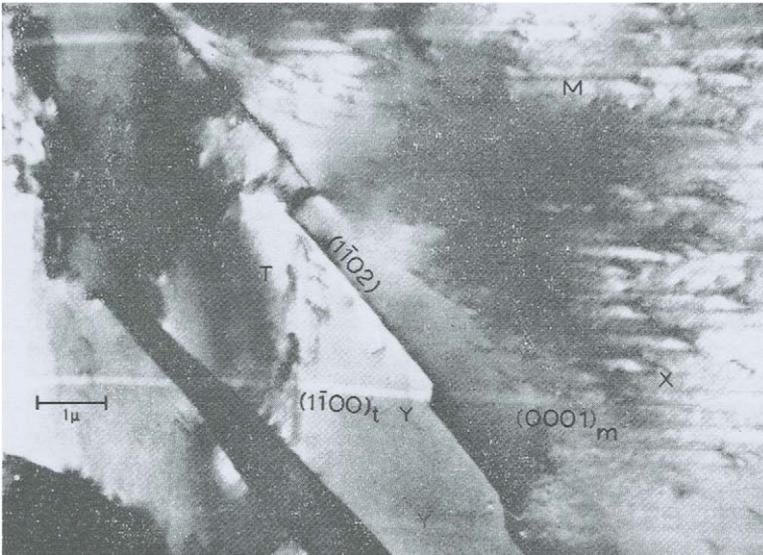


FIG. 20.32. Electron micrograph showing slip transferred from the matrix basal plane (0001) to the  $\{1\bar{1}00\}$  plane of a  $\{1\bar{1}02\}$  barrier twin in zinc (courtesy of Tomsett and Bevis, 1969).

vector. General conditions were first given by Cahn (1953a) who showed that the traces of the  $K_1$  planes matrix- $A$  and  $B-C$  in that of matrix- $B$  must be parallel and the direction, sense and magnitude of shear for matrix- $A$  and  $B-C$  twinning must be the same. He considered specifically type I or compound twinning in which the  $K_1$  planes of  $A$  and  $B$  intersect in the  $\eta_1$  direction common to both, so that  $B$  undergoes no displacement normal to its  $K_1$  plane. This means that the twinning dislocations of  $A$  meet the  $A-B$  interface in screw orientation and can cross-slip into  $B$  without leaving a defect at the interface. The  $K_1$  planes of  $A$  and  $C$  are then mirror images in that of  $B$ . In an alternative geometry pointed out by Liu (1963), the  $K_1$  planes of  $A$  and  $C$  are now parallel, so that the crossing twin  $A$  is undeviated but  $B$  is displaced. Individual twinning dislocations of the matrix- $A$  interface each produce a step as they cross into  $B$  and the successive steps rotate the  $B-C$  interface away from the matrix- $B$  interface.

Many studies have been made of the propagation of slip lines or deformation twins across existing coherent matrix-twin interfaces in f.c.c., b.c.c. and h.c.p. materials. Among more routine results is the suggestion of Rowlands *et al.* (1968) (see p. 891) that an unusual {5, 8, 11} twinning mode found in some b.c.c. and b.c.t. ferrous martensites arises because such twins may propagate undeviated across the closely spaced {112} transformation twin array in the martensite, at least in cubic alloys. Another interesting case arises if the deformation in a twin crossed by another is produced not by formation of a secondary twin  $C$  but by an equivalent slip shear. This allows relaxation of Cahn's conditions and has been studied in a f.c.c. cobalt iron alloy by Mahajan and Chin (1973a).

#### REFERENCES

- ABELL, J. S., CROCKER, A. G., and GUYONCOURT, D. M. M. (1971) *J. Mater. Sci.* **6**, 361.  
 ALLEN, N. P., HOPKINS, B. E., and MCLENNAN, J. E. (1956) *Proc. R. Soc. London, Ser. A* **234**, 221.  
 ANDRADE, E. N. da C. and HUTCHINGS, P. J. (1935) *Proc. R. Soc. London, Ser. A* **148**, 120.  
 ANTONOPOULOS, J. G., KARAKOSTAS, Th., KOMNINOY, Ph., and DELAVIGNETTE, P. (1989) *Acta Metall.* **36**, 2483.  
 ARUNCHALAMAN, V. S. and SARGENT, G. M. (1971) *Acta Metall.* **5**, 949.  
 ASHBY, M. F. (1972) *Surf. Sci.* **31**, 498.  
 BACON, D. J. and SERRA, A. (1994) *Twinning in Advanced Materials*, eds. M. H. Yoo and M. Wuttig, p. 275, Metallurgical Society of AIME, Warrendale, PA.  
 BALLUFFI, R. W., KOMEN, Y., and SCHÖBER, T. (1972) *Surf. Sci.* **31**, 68.  
 BELL, R. L. and CAHN, R. W. (1953) *Acta Metall.* **1**, 752.  
 BEVIS, M. and CROCKER, A. G. (1968) *Proc. R. Soc. London, Ser. A* **304**, 123; (1969) *Ibid.* **313**, 509.  
 BEVIS, M. and VITEK, V. (1970) *Scr. Metall.* **4**, 719.  
 BIGGS, W. D. and PRATT, P. L. (1958) *Acta Metall.* **6**, 694.  
 BILBY, B. A. (1953) *Year Book of the Royal Society*, p. 217, The Royal Society, London.  
 BILBY, B. A. and CHRISTIAN, J. W. (1956) *The Mechanism of Phase Transformations in Metals*, p. 121, The Institute of Metals, London.  
 BILBY, B. A. and CROCKER, A. G. (1965) *Proc. R. Soc. London, Ser. A* **288**, 240.  
 BIRNBAUM, H. and READ, T. A. (1960) *Trans. Metall. Soc. AIME* **218**, 662.  
 BLEWITT, T. H., COLTMAN, R. R., and REDMAN, J. K. (1957) *J. Appl. Phys.* **28**, 651.  
 BOLLING, G. F. and RICHMAN, R. H. (1967) *Can. J. Phys.* **45**, 541.  
 BOLLMANN, W. (1970) *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin.  
 BOUCHER, N. A. and CHRISTIAN, J. W. (1972) *Acta Metall.* **20**, 581.

- BRAISAZ, T., BACON, D. J., NOUET, G., POND, R. C., RUTERANA, P., and SERRA, A. (1996) *Mater. Sci. Forum* **207**.
- BRISTOWE, P. D. and CROCKER, A. G. (1975) *Philos. Mag.* **31**, 803; (1976) *Ibid.* **33**, 357; (1977) *Acta Metall.* **25**, 1363.
- BULLOUGH, R. (1957) *Proc. R. Soc. London, Ser. A* **241**, 568.
- CAHN, J. W. (1977) *Acta Metall.* **25**, 1021.
- CAHN, R. W. (1953a) *Acta Metall.* **1**, 49; (1953b) *Nuovo Cimento, Suppl.* **10**, 350.
- CAHN, R. W. and COLL, J. A. (1961) *Acta Metall.* **9**, 138.
- CHAKRABORTY, S. B. and STARKE, E. A., Jr. (1975) *Acta Metall.* **23**, 63.
- CHIN, G. Y., HOSFORD, W. F., and MENDORE, D. R. (1969) *Proc. R. Soc. London, Ser. A* **309**, 433.
- CHRISTIAN, J. W. (1959) *Acta Metall.* **7**, 218; (1969a) *Comments Solid State Phys.* **2**, 50; (1969b) *Physics of Strength and Plasticity*, ed. A.S. Argon, p. 85, MIT Press, Boston, MA; (1970) *Proc. 2nd Int. Conf. Strength of Metals and Alloys*, p. 29, American Society for Metals, Metals Park, OH; (1983) *Metall. Trans.* **14A**, 1237.
- CHRISTIAN, J. W. and CROCKER, A. G. (1980) *Dislocations and Solids*, Vol. 3, ed. F. R. N. Nabarro, p. 165, North-Holland, Amsterdam.
- CHRISTIAN, J. W. and KNOWLES, K. M. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 1185, Metallurgical Society of AIME, Warrendale, PA.
- CHRISTIAN, J. W. and LAUGHLIN, D. E. (1987) *Scr. Metall.* **21**, 1131; (1988) *Acta Metall.* **28**, 1617.
- CHRISTIAN, J. W. and MAHAJAN, S. (1995) *Prog. Mater. Sci.* **39**, 1.
- CHRISTIAN, J. W., MASTERS, B. C., MITCHELL, A., and PARGETER, F. W. J. (1960) *Acta Crystallogr.* **13**, 114.
- CHURCHMAN, A. T., GEACH, G. A., and WINTON, J. (1956) *Proc. R. Soc. London, Ser. A* **238**, 194.
- COTTELL, A. H. and BILBY, B. A. (1951) *Philos. Mag.* **42**, 573.
- COULING, S. L., PASHAK, J. F., and STURKEY, L. (1959) *Trans. Am. Soc. Met.* **51**, 94.
- COULING, S. L. and ROBERTS, C. S. (1956) *Acta Crystallogr.* **9**, 972.
- COX, J. J., HORNE, G. T., and MEHL, R. F. (1957) *Trans. Am. Soc. Met.* **44**, 118.
- CROCKER, A. G. (1959) *Ph.D. Thesis*, Sheffield University, Sheffield, UK; (1962) *Philos. Mag.* **7**, 1901; (1965) *J. Nucl. Mater.* **16**, 306; (1971) *Ibid.* **41**, 167.
- CROCKER, A. G. and BEVIS, M. (1970) *The Science, Technology and Application of Titanium*, eds. R. Jaffee and N. Promisel, p. 453, Pergamon Press, Oxford.
- CROCKER, A. G., HECKSCHER, R., BEVIS, M., and GUYONCOURT, D. M. M. (1966) *Philos. Mag.* **13**, 1191.
- DELAMOTTE, E. and ALSTETTER, C. (1969) *Trans. Metall. Soc. AIME* **245**, 651.
- FEARON, E. O. and BEVIS, M. (1974) *Acta Metall.* **22**, 991.
- FRANK, F. C. (1953) *Acta Metall.* **1**, 71.
- FRANK, F. C. and STROIL, A. N. (1952) *Proc. Phys. Soc. London, Ser. B* **65**, 295.
- FRANK, F. C. and THOMPSON, N. (1955) *Acta Metall.* **3**, 30.
- FREISE, E. J. and KELLY, A. (1961) *Proc. R. Soc. London, Ser. A* **264**, 269.
- FUJITA, H. and MORI, T. (1975) *Scr. Metall.* **9**, 631.
- GALLAGHER, P. C. J. (1966a) *J. Appl. Phys.* **37**, 1719; (1966b) *Phys. Status Solidi* **126**, 95.
- GEROLD, V. and KARNTHALER, H. P. (1989) *Acta Metall.* **37**, 2177.
- GOO, E., DUERIG, T., MELTON, K., and SINCLAIR, R. (1985) *Acta Metall.* **33**, 1725.
- GRATIAS, D. and THALAL, A. (1988) *Philos. Mag. Lett.* **57**, 62.
- GREEN, M. L. and COHEN, M. (1979) *Acta Metall.* **27**, 1523.
- GUEDO, J. Y. and RIEU, J. V. (1978) *Scr. Metall.* **12**, 927.
- GUIMIER, A. and STRUDEN, J. L. (1970) *Proc. 2nd Int. Conf. Strength of Metals and Alloys*, p. 1145, American Society for Metals, Metals Park, OH.
- GUYONCOURT, D. M. M. and CROCKER, A. G. (1968) *Acta Metall.* **16**, 523.
- HAASEN, P. (1958) *Philos. Mag.* **3**, 384.
- HAASEN, P. and KING, A. H. (1960) *Z. Metallkd.* **51**, 700.
- HALL, E. O. (1954) *Twinning*, Butterworths, London.
- HANSSON, B. and BARNES, R. S. (1965) *Acta Metall.* **12**, 315.
- HARNECKER, K. and RASSOW, E. (1924) *Z. Metallkd.* **16**, 312.

- HARRIS, W. F. (1970) *Philos. Mag.* **22**, 949.
- HIRTH, J. P. (1963) *Deformation Twinning*, eds. R. E. Reed-Hill, J. P. Hirth and H. C. Rogers, p. 112, Gordon and Breach, New York.
- HIRTH, J. P. and LOTHE, J. (1982) *Theory of Dislocations*, 2nd edn., p. 818, McGraw-Hill, New York.
- HOLDEN, J. (1952) *Philos. Mag.* **43**, 976.
- HONG, S. I. and LAIRD, C. (1990) *Acta Metall.* **38**, 1581.
- HULL, D. (1960) *Acta Metall.* **8**, 11.
- JASWON, M. A. and DOVE, D. B. (1956) *Acta Crystallogr.* **9**, 621; (1957) *Ibid.* **10**, 14; (1960) *Ibid.* **13**, 232.
- KEAR, B. H. and OBLAK, J. M. (1974) *Dissociation des Dislocations* (Supplement to J. Phys.(Paris), Colloq.) **35-C7**, 35.
- KELLY, A. (1952) *Proc. Phys. Soc. London, Ser.A* **67**, 403.
- KIHO, H. (1954) *J. Phys. Soc. Jpn.* **9**, 739; (1958) *Ibid.* **13**, 269.
- KRONBERG, M. L. (1959) *J. Nucl. Mater.* **1**, 85; (1961) *Acta Metall.* **9**, 970.
- LAVES, F. (1952) *Naturwissenschaften* **39**, 456; (1966) *Acta Metall.* **14**, 58.
- LAY, S. and NOUET, G. (1994) *Philos. Mag.* **70A**, 261.
- LEVINE, P. and STEINHARDT, P. J. (1986) *Phys. Rev.B* **34**, 596.
- LIU, Y. C. (1963) *Trans. Metall. Soc. AIME* **227**, 775.
- LLOYD, L. T. and CHISWICK, H. H. (1955) *Trans. Metall. Soc. AIME* **203**, 1206.
- MAGEE, C. L., HOFFMAN, D. W., and DAVIES, R. G. (1971) *Philos. Mag.* **23**, 1531.
- MAHAJAN, S. (1971) *Philos. Mag.* **23**, 781; (1972a) *Ibid.* **26**, 161; (1975) *Acta Metall.* **23**, 671.
- MAHAJAN, S., BARRY, D. E., and EYRE, B. L. (1970) *Philos. Mag.* **21**, 43.
- MAHAJAN, S. and CHIN, G. Y. (1973a) *Acta Metall.* **21**, 1353.
- MAHAJAN, S. and WILLIAMS, D. F. (1973a) *Int. Metall. Rev.* **18**, 43.
- MATHEWSON, C. H. and EDMUNDS, G. H. (1928) *Trans. Metall. Soc. AIME* **80**, 311.
- MIKKOLA, D. E. and COHEN, J. B. (1966) *Acta Metall.* **14**, 105.
- MILLARD, D. J. and THOMPSON, N. (1952) *Philos. Mag.* **43**, 422.
- MINONISHI, V., ISHIOKA, S., KOIWA M., and MOROZUMI, S. (1982) *Phys. Status Solidi A* **71**, 253.
- MIURA, S., TAKAMURA, J., and NARITA, N. (1968) *Proc. Int. Conf. Strength of Metals and Alloys* (Supplement to *Trans. Jpn. Inst. Met.*) **9**, 555.
- MOORE, A. J. W. (1952) *Proc. Phys. Soc. London, Ser.B* **55**, 956; (1955) *Acta Metall.* **3**, 163.
- MORI, T. and FUJITA, H. (1977) *Trans. Jpn. Inst. Met.* **18**, 17; (1980) *Acta Metall.* **28**, 771.
- MORI, T., FUJITA, H., and TAKEMORI, S. (1981) *Philos. Mag.* **44**, 1277.
- NABARRO, F. R. N. (1952) *Adv. Phys.* **1**, 269.
- NARITA, N. and TAKAMURA, J. (1974) *Philos. Mag.* **29**, 1001.
- NESBITT, L. A. and LAUGHLIN, D. E. (1980) *Acta Metall.* **28**, 989.
- OGAWA, K. (1965) *Philos. Mag.* **11**, 217.
- OLIVER, D. S. (1952) *Research(London)* **5**, 45.
- OLSON, G. B. (1981) *Acta Metall.* **29**, 1475.
- OLSON, G.B. and COHEN, M. (1979) *Acta Metall.* **27**, 1007.
- OOKAWA, A. (1957) *J. Phys. Soc. Jpn.* **12**, 825.
- OROWAN, E. (1954) *Dislocations in Metals*, ed. M. Cohen, p. 116, AIME, New York.
- PASHLEY, D. W., ROBERTSON, J. L., and STOWELL, M. J. (1969) *Philos. Mag.* **19A**, 83.
- PAXTON, A. T. (1994) *Twinning in Advanced Materials*, eds. M. H. Yoo and M. Wuttig, p. 27, Metallurgical Society of AIME, Warrendale, PA.
- PAXTON, H. W. (1953) *Acta Metall.* **1**, 141.
- PEISSKER, E. (1965) *Z. Metallkd.* **56**, 155.
- PIROUZ, P. (1987) *Scr. Metall.* **21**, 1463; (1989) *Ibid.* **23**, 401.
- PIROUZ, P. and HAZZLEDINE, P. M. (1991) *Scr. Metall.* **25**, 1167.
- POND, R. C. (1985) *Dislocations and Properties of Real Materials*, p. 71, The Institute of Metals, London; (1988) *Dislocations in Solids*, Vol. 8, ed. F. R. N. Nabarro, p. 1, North-Holland, Amsterdam; (1995) *Interface Sci. (Netherlands)* **2**, 299.
- POND, R. C. and BOLLMANN, W. (1979) *Philos. Trans. R. Soc. London, Ser.A* **292**, 449.
- POND, R. C. and GARCIA-GARCIA, L. M. F. (1981) *EMAG, Inst. Phys. Conf. Ser. No. 61*, p. 495. The Institute of Physics London.

- POND, R. C. and VLACHAVAS, D. S. (1983) *Proc. R. Soc. London, Ser. A* **386**, 95.
- PRATT, P. L. and PUGH, S. F. (1952) *J. Inst. Met.* **80**, 653.
- PRICE, P. B. (1960) *Philos. Mag.* **5**, 873; (1961) *Proc. R. Soc. London, Ser. A* **260**, 251; (1962) *J. Appl. Phys.* **32**, 1746.
- RAPPERPORT, E. J. (1959) *Acta Metall.* **7**, 254.
- RAPPERPORT, E. J. and HARTLEY, C. S. (1960) *Trans. Metall. Soc. AIME* **218**, 869.
- RECHTIEN, J. J., CROCKER, A. G., and NELSON, R. D. (1971) *J. Nucl. Mater.* **40**, 134.
- REED-HILL, R. E. (1960) *Trans. Metall. Soc. AIME* **218**, 554.
- REED-HILL, R. E., HIRTH, J. P., and ROGERS, H. C. (eds.) (1964) *Deformation Twinning*, Gordon and Breach, New York.
- REED-HILL, R. E. and ROBERTSON, W. D. (1957) *Acta Metall.* **5**, 717.
- REED-HILL, R. E., SLIPPY, W. A., Jr., and BUTEAU, L. J. (1963) *Trans. Metall. Soc. AIME* **227**, 976.
- REID, C. A., GILBERT, A., and HAHN, G. T. (1966) *Acta Metall.* **14**, 975.
- RICHARDS, C. E. and REID, C. N. (1968) *Trans. Metall. Soc. AIME* **242**, 1831.
- RICHARDS, M. J. and CAHIN, J. W. (1971) *Acta Metall.* **19**, 1263.
- RICHMAN, R. H. (1964) *Deformation Twinning*, eds. R. E. Reed-Hill, J. P. Hirth and H. C. Rogers, p. 237, Gordon and Breach, New York.
- RICHMAN, R. H. and CONARD, G. P. (1963) *Trans. Metall. Soc. AIME* **227**, 774.
- ROSENBAUM, H. S. (1961) *Acta Metall.* **9**, 742; (1964) *Deformation Twinning*, eds. R. E. Reed-Hill, J. P. Hirth and H. C. Rogers, p. 43, Gordon and Breach, New York.
- ROWLANDS, P. C., FEARON, E. O., and BEVIS, M. (1968) *Trans. Metall. Soc. AIME* **242**, 1559; (1970) *J. Mater. Sci.* **5**, 169.
- SAXL, I. (1968) *Czech. J. Phys.* **B18**, 39.
- SEGER, A. (1956) *Z. Metallkd.* **47**, 653.
- SERRA, A. and BACON, D. J. (1986) *Philos. Mag.* **54A**, 793.
- SERRA, A., BACON, D. J., and POND, R. C. (1988) *Acta Metall.* **36**, 3183.
- SIECHTMAN, D., BLACKBURN, M. J., and LIPSITT, H. A. (1974) *Metall. Trans.* **5**, 1373.
- SLEESWYK, A. W. (1962) *Acta Metall.* **10**, 803; (1963) *Philos. Mag.* **8**, 1467; (1974) *Ibid.* **29**, 407.
- SMITH, E. (1960) Personal communication.
- SMITH, J. H. and GAUNT, P. (1961) *Acta Metall.* **9**, 819.
- SOKURSKII, Yu. N. and PROTSENKO, L. N. (1958) *Sov. J. At. Energy* **4**, 579.
- SPRIET, B. (1964) *J. Nucl. Mater.* **12**, 159.
- SREE HARSHA, K. S. (1981) *Metall. Trans.* **12A**, 365.
- STEEDS, J. W. and HAZZLEDINE, P. M. (1964) *Discuss. Faraday Soc.* **38**, 103.
- STOLOFF, N. S. and GENSAMER, M. (1963) *Trans. Metall. Soc. AIME* **224**, 70.
- SUMINO, K. (1966) *Acta Metall.* **14**, 1607.
- SUTTON, A. P. (1988) *Acta Metall.* **36**, 1291.
- SUTTON, A. P. and CHRISTIAN, J. W. (1982) *J. Phys. (Paris), Colloq.* **12-C4**, 197.
- SUTTON, A. P. and VITEK, V. (1983) *Philos. Trans. R. Soc. London, Ser. A* **309**, 1, 37, 55.
- SUZUKI, H. and BARRETT, C. S. (1958) *Acta Metall.* **6**, 156.
- TAWANCY, H. M. (1981) *J. Mater. Sci.* **16**, 2883.
- THORNTON, P. R. and MITCHELL, T. E. (1962) *Philos. Mag.* **7**, 361.
- TOMSETT, D. I. and BEVIS, M. (1969) *Philos. Mag.* **19**, 129, 533.
- VAIDYA, S. and MAHAJAN, S. (1980) *Acta Metall.* **28**, 1123.
- VANDERSHAEVE, G. and SARRAZIN, T. (1977) *Phys. Status Solidi* **A43**, 459.
- VANDERSHAEVE, G., SARRAZIN, T., and ESCAIG, B. (1979) *Acta Metall.* **27**, 1251.
- VENABLES, J. A. (1960) *Proc. Eur. Conf. Electron Microscopy, Delft, The Netherlands*, Vol. 1, p. 443; (1961) *Philos. Mag.* **6**, 379; (1974) *Ibid.* **30**, 1165.
- VITEK, V. (1968) *Philos. Mag.* **18**, 773; (1985) *Dislocations and Properties of Real Materials*, p. 30, The Institute of Metals, London.
- VOHRINGER, O. (1970) *Proc. 2nd Int. Conf. Strength of Metals and Alloys*, p. 294, American Society of Metals, Metals Park, OH.
- VOTAVA, E. and SLEESWYK, A. W. (1962) *Acta Metall.* **10**, 965.
- WESTLAKE, D. G. (1961) *Acta Metall.* **9**, 327; (1963) *Deformation Twinning*, eds. R. E. Reed-Hill, J. P. Hirth and H. C. Rogers, p. 29, Gordon and Breach, New York.

- WHELAN, M. J. (1958) *Proc. R. Soc. London, Ser. A* **249**, 114.  
 YAMAGUCHI, M., UMAKOSHI, Y., and YAMANE, T. (1987a) *Philos. Mag.* **55A**, 301; (1987b) *High Temperature Metallic Alloys*, eds. N. Stoloff, C. C. Kock, C. T. Liu and O. Izumi, p. 275, MRS Symposium Press, Pittsburgh, PA.  
 YAMAGUCHI, M. and VITEK, V. (1976) *Philos. Mag.* **34**, 1.  
 YOO, M. H. (1981) *Metall. Trans.* **12A**, 409.  
 YOO, M. H. and WEI, C. T. (1966) *Philos. Mag.* **14**, 573.

#### FURTHER REFERENCES

- ALTSCHULER, T. L. and CHRISTIAN, J. W. (1966) *Acta Met.* **14**, 903.  
 ANDROUSSI, Y., VANDERSCHAEVE, G. and LIEFEBVRE, A. (1989) *Phil. Mag.* **A59**, 1189.  
 BOIVIN, P., RABIER, J. and GAREM, H. (1990b) *Phil. Mag.* **A61**, 647.  
 BOIVIN, P., RABIER, J. and GAREM, H. (1990a) *Phil. Mag.* **A61**, 619.  
 CHIN, G. Y., HOSFORD, W. F. and MENDORE, D. R. (1969) *Proc. Roy. Soc.* **A309**, 433.  
 EDMONSON, B. (1961) *Proc. Roy. Soc.* **A264**, 176.  
 GOUGH, H. J. (1933) *Proc. A.S.T.M.* **33**, 3.  
 GRAY III, G. T. (1988) *Acta Met.* **36**, 1745.  
 HONDA, R. (1961) *Proc. Phys. Soc. Jap.* **16**, 1309.  
 HULL, D. (1961) *Acta Met.* **9**, 909.  
 HULL, D. (1963a) *Fracture of Solids*, p. 147. Interscience, NY.  
 HULL, D. (1963b) *Proc. Roy. Soc.* **A274**, 5.  
 HULL, D. (1963c) *Deformation Twinning* (edited by R. E. Reed-Hill, J. P. Hirth and H. C. Rogers), p. 121, Gordon and Breach, New York.  
 LAVRENTYEV, F. F., SALITA, O. P. and KAZAROV, YU. G. (1968) *Fiz. Met. Metallo.* **26**, 438.  
 LE, T. D., BERNSTEIN, I. M. and MAHAJAN, S. (1993) *Acta Met. et Mater.* **41**, 3363.  
 LESLIE, W. C., HORNBOGEN, E. and DIETER, W. E. (1962) *J. Iron Steel Inst.* **200**, 622.  
 MAHAJAN, S. (1969) *Phys. Stat. Sol.* **33**, 291.  
 MAHAJAN, S. (1970) *Phys. Stat. Sol.* **A2**, 217.  
 MAHAJAN, S. (1972b) *J. Phys. F: Metal Physics* **2**, 19.  
 MAHAJAN, S. and BARILETTI, A. F. (1971) *Acta Met.* **19**, 1111.  
 MAHAJAN, S. and CHIN, G. Y. (1973b) *Acta Met.* **21**, 173.  
 MAHAJAN, S. and CHIN, G. Y. (1974) *Acta Met.* **22**, 1113.  
 MCHARGUE, C. J. (1964) *Trans. Met. Soc. Amer. Inst. Min. (Metall.) Eng.* **224**, 334.  
 ODINKOVA, L. P. (1967) *Rus. Metall. (Metally)* **1**, 66.  
 PARTRIDGE, P. G. (1967) *Met. Rev.* **12**, 169.  
 PATON, N. E. and BACKOFEN, W. A. (1970) *Metall. Trans.* **1**, 2839.  
 PIROUZ, P., CHAIM, R., DAHMEN, U. and WESTMACOTT (1990) *Acta Met.* **313**.  
 PRIESTNER, R. and LESLIE, W. C. (1965) *Phil. Mag.* **11**, 895.  
 RABIER, J. and BOIVIN, P. (1990) *Phil. Mag.* **A61**, 673.  
 REED-HILL, R. E. and DAHLBERG, E. P. (1966) *Electrochem. Technol.* **4**, 303.  
 RICHMAN, R. H. (1963) *Deformation Twinning* (edited by R. E. Reed-Hill, J. P. Hirth and H. C. Rogers), p. 237, Gordon and Breach, New York.  
 RICHMAN, R. H. and CONRAD, G. P. (1963) *Trans. Met. Soc. Amer. Inst. Min. (Metall.) Eng.* **227**, 774.  
 ROBERTSON, I. M. (1986) *Phil. Mag.* **A54**, 821.  
 ROSENFELD, A. R., AVERBACH, B. L. and COHEN, M. (1963) *Acta Met.*, **11**, 1100.  
 SERRA, A., POND, R. C. and BACON, D. J. (1991) *Acta Metall. Mater.* **39**, 1469.  
 TERASAKI, F. (1967) *Acta Met.* **15**, 1057.  
 TU, K. N. and TURNBULL, D. (1970) *Acta Met.* **18**, 915.  
 WESSEL, E. T., FRANCE, L. L. and BEGLEY, R. T. (1961) *Columbian Metallurgy*, p. 485, Interscience, NY.  
 ZUKAS, E. G. and FOWLER, C. M. (1961) *Response of Metals to High Velocity Deformation*, p. 343, Interscience, NY.

## CHAPTER 21

# *Characteristics of Martensitic Transformations*

### 89. ATOMIC MOVEMENTS IN THE TRANSFORMATION

Although it was stated in Chapter 1 that the experimental criterion most useful for defining a martensitic transformation is the observation of a shape change in a transforming region, this is now recognized to be not universally true. As pointed out in Chapter 16, some transformations involving diffusion may nevertheless have a coherent interface and a lattice correspondence which persists as the new phase grows. However, a shape change does imply a lattice correspondence and in most cases this means a martensitic transformation.

The division of heterogeneous reactions in the solid state into two groups arose gradually from the early work on the martensitic reaction in quenched steels, but a definitive list of the characteristics of the two kinds of transformation was first given by Troiano and Greninger (1946). Later papers and review articles gradually clarified the nature of the martensitic transformation, but the characteristics listed by Troiano and Greninger continued to be used as the basis of the division. In the meantime, however, work by Kurdjumov (1948) and by Kurdjumov and Maximova (1948, 1950) had shown that particular kinetic characteristics were absent in certain transformations. Although Kurdjumov's results were initially greeted with considerable scepticism, they were soon confirmed by American workers and it became evident that a more precise definition of a martensitic transformation was required. The suggestion that such a definition be based on a change of shape of the transformed region was made by Hull (1954) and by Bilby and Christian (1956). Despite the possibility of diffusional displacive transformations (see pp. 757-759) the change of shape remains the experimental criterion most readily associated with martensitic transformation.

The change of shape which produces tilting of an originally flat polished surface implies that the product region remains highly coherent with the parent. In contrast, such coherence is present only in the very early stages of most nucleation and growth transformations. As discussed in detail in Section 52, the strain energy produced by a coherent transformation is very much larger than that of an incoherent transformation, and this usually leads to loss of coherency ("breaking away") when the particle of new phase becomes sufficiently large for its surface free energy to be negligible. In a martensitic reaction, the extra strain energy is tolerated because some degree of coherency is essential

to the mechanism of transformation, and alternative forms of transformation with smaller strain energies do not take place because their rates are very much slower. The martensitic transformation thus occurs because of the existence of an easy growth mechanism, not requiring atomic diffusion, which leads to the rapid production of a new phase and a net lowering of the free energy. The question of growth mechanism is the central feature of a martensitic change and emphasizes the importance of the martensitic interface in such a reaction.

If atomic interchange does not play a significant role in the mechanism of transformation, it follows that the martensite must have the same chemical composition as the parent phase. The absence of atomic interchange is confirmed by the growth of martensite at very low temperatures and by the fact that ordered phases remain ordered after transformation. One significant exception to the constant composition rule is the formation of bainite in steels. This takes place in a temperature range where the iron atoms and any substitutional solutes are completely immobile, but where the interstitial carbon atoms have appreciable mobility. The iron atoms then undergo a displacive (martensitic type) transformation whilst the carbon atoms diffuse to lower the carbon content of the supersaturated bainitic ferrite. This mechanism, suggested by Ko and Cottrell (1952), is clearly confined to systems with interstitial solutes.

In the absence of diffusion, the atomic movements must be orderly and coordinated, and the neighbours of most atoms remain unchanged. A set of atoms defining a space lattice will define a new space lattice if given displacements equivalent to a homogeneous deformation and, if atomic interchange or displacements over distances greater than interatomic separations are forbidden, it follows that a change of lattice can be produced only by such a homogeneous deformation, and this will be called the lattice deformation. Unless the lattice deformation leaves one plane undistorted, i.e. unchanged in the magnitude and relative orientations of all vectors within the plane, it is not possible to have an invariant plane; however, if one undistorted plane exists, so does a second undistorted plane and either of these may become an invariant plane by adding a suitable rotation to the description. The set of lattice points related by a lattice deformation may comprise only a fraction of the lattice points of the original structure, but it is evident that a change of structure without atomic interchange can be accomplished only if *some* unit cell of the parent structure is effectively deformed into a corresponding unit cell of the product. The atoms within this unit cell then undergo additional displacements ("shuffles") to complete the structural change.

The shape deformation of a martensite plate always tilts the free surface about the line of intersection of the plate and the surface, and this is characteristic of an invariant plane strain. Application of this measured deformation to the parent structure, however, will not, in general, give the product structure, because as already stated two lattices cannot, in general, be connected by an invariant plane strain. As will be discussed in the next chapter, a satisfactory physical picture is obtained if the lattice deformation  $S$  is homogeneous only over small volumes of the product phase, and the deformations in different regions combine together to give the observed shape deformation. The structural change is thus given by a homogeneous lattice deformation which may be the same in

adjacent small regions without, however, a common origin. The regions are displaced from one another by a lattice vector. The periodic displacements in the interface produce a simple shear which restores the lattice as the interface is displaced, and the macroscopic or shape deformation is the result of the combined lattice deformation and "lattice-invariant deformation". The above description distinguishes between the deformation of the crystal lattice and that of large vectors crossing many unit cells. Figure 21.1 shows schematically the types of change which may occur. The homogeneous deformation of one lattice into another is represented in Fig. 21.1(b) and this corresponds to a shape change which is identical to that of a unit cell. A change of shape without changing the lattice may be produced by slip deformation [Fig. 21.1(c)] or more generally by adding or removing lattice sites; this is the lattice-invariant transformation. By combining a lattice deformation with a lattice-invariant deformation, a zero net shape change is produced [Fig. 21.1(d)].

Figure 21.1 is drawn on the assumption that the slip vector is a lattice vector so that the lattice-invariant deformation produces no defects in the structure. In addition, it may be possible to modify the shape change by translating parts of the structure through displacements which represent partial dislocations of the product lattice. This may also loosely be described as a lattice-invariant deformation as the unit cells are unchanged, but there will now be surfaces of discontinuity (stacking faults) within the product. This type of deformation could not occur unless the faults have very low energies, and the slip vectors are confined to those which bound low energy faults.

It is also possible to vary the shape deformation associated with a given lattice change by having different lattice deformations alternating on a fine scale throughout the transformed region. The alternating regions of martensite will then have different orientations, and there will be real interfaces inside the product instead of the slip planes of Fig. 21.1(d). This configuration is possible only if the interfaces have very low energies,

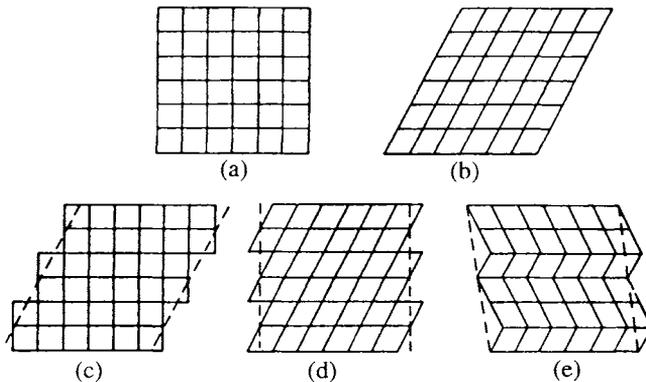


FIG. 21.1. Types of deformation in martensitic transformation (after Bilby and Christian, 1956): (a) original undeformed crystal block; (b) lattice deformation; (c) lattice-invariant (or dislocation) deformation by slip; (d) lattice deformation and lattice-invariant deformation combining to give zero total shape deformation; (e) shape deformation caused by varying lattice deformation in different regions.

and this condition is likely to be met only by coherent  $K_1$  interfaces, giving a product which is a stack of fine twins. Figure 21.1(e) illustrates this possibility.

It is useful to emphasize here that a martensitic transformation cannot usually be described as a lattice deformation which is uniform throughout the transformed volume. The interface must be an invariant plane of the total shape deformation, and this is achieved in one of the ways just described. The macroscopic shape deformation is the net result of the individual lattice deformations and of any lattice-invariant deformation.

The theoretical conclusion about the nature of the shape deformation is supported by a mass of experimental evidence. Crystals formed by a martensitic transformation are visible on a surface pre-polished before transformation because of the tilting already mentioned. Similarly, straight scratches before transformation are deviated into straight lines of different orientation within the product regions as shown in Fig. 21.2. The shape change must thus be as shown in Fig. 21.3. Consider the parallel lines  $AB$  and  $CD$  of this figure.

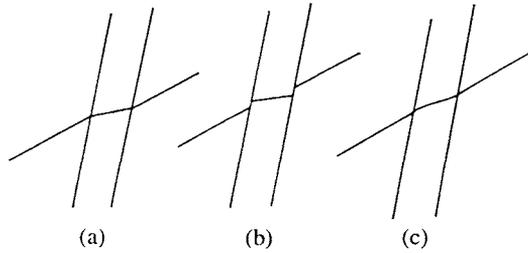


FIG. 21.2. Possible distortions of straight reference mark produced by martensite plate: (a) shape deformation is an invariant plane strain; (b) loss of coherency at interface; (c) elastic distortion in habit plane.

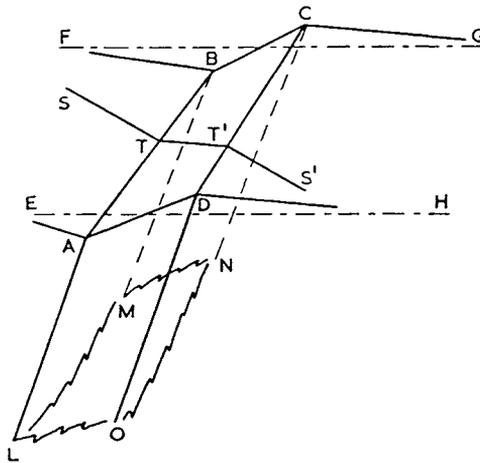


FIG. 21.3. The shape deformation produced by a martensite plate (after Bilby and Christian, 1956). The formation of a martensite plate  $ABCDOLMN$  produces tilts along  $AB$  and  $DC$  where the habit planes intersect the free surface, and deviates a straight scratch into  $STT'S'$ .

They both lie in the habit plane and are invariant (unchanged in magnitude and direction) during the formation of the plate. Both are arbitrary directions in the interface, being fixed by the orientation of the free surface, so that it follows that any arbitrary line lying in the plane of the plate is invariant in the transformation. The whole habit plane thus consists of a set of invariant vectors and hence is an invariant plane. Thus *experimental observation* shows that the habit plane is an invariant plane of the shape deformation, which is the essential assumption of the crystallographic theories.

The significance of these experimental results was realized at least partly by many of the early workers but was first stated explicitly by Bowles and Mackenzie (1954). The macroscopic nature of the observations allows only the deduction that the habit plane is approximately an invariant plane of the shape deformation, and small effects of the type shown in Fig. 21.2(c) might not be detectable. Bowles and Mackenzie assumed in fact that small changes in length (up to about 1%) in the habit plane are possible, but that all vectors retain their original orientation. Thus they assumed that the shape deformation is an invariant plane strain plus (possibly) a small uniform dilatation or contraction. Once some deviation from the macroscopic matching condition is allowed, however, there seems no reason why this should be uniform in all directions. Some theories of martensite crystallography have been developed with the explicit or implicit assumption that macroscopic vectors in the habit plane undergo small changes in length during transformation, the percentage change varying with the direction of the vector.

Within the limits of the observations already discussed, it is very difficult to determine experimentally whether or not there is any deviation from the invariant plane strain condition for the habit plane. The only feasible procedures seem to be the use of surface replicas or the indirect method of comparison of observed and predicted crystallographic parameters of the transformation (see Chapter 22), some of which are quite sensitive to the actual situation. There is the further difficulty that a considerable scatter is often found among the measured habit planes in a single specimen, not all of which can readily be attributed to experimental error (Otte and Read, 1957). This may indicate that the details of the structure vary from plate to plate. Individual plates, especially lenticular plates enclosed in the matrix, may encounter different local stress situations which can modify the crystallography. In the few transformations where a planar martensite interface traverses the whole cross-section of a single crystal, the habit plane appears to be accurately an invariant plane of the shape deformation, there being excellent agreement between predicted and observed parameters.

The simplest martensitic transformations are those between structures which may be regarded as layers of close-packed spherical atoms in various stacking sequences to give structures such as f.c.c., h.c.p., d.h.c.p., 9R, 18R, etc. For atoms of almost the same size, fully coherent interfaces can form parallel to the hexagonal layers. The transformation from the high temperature f.c.c. phase to the low temperature h.c.p. phase which occurs in cobalt and its alloys, in iron–manganese alloys and in certain alloy steels has been much investigated (Takeuchi and Honma, 1957; Bibring *et al.*, 1958–9; Gaunt and Christian, 1959a; Hayzelden *et al.*, 1991; Waitz and Karnthaler, 1996). If the h.c.p. structure has ideal axial ratio, the transformation may be effected by a simple shear of magnitude  $8^{-1/2}$  on a

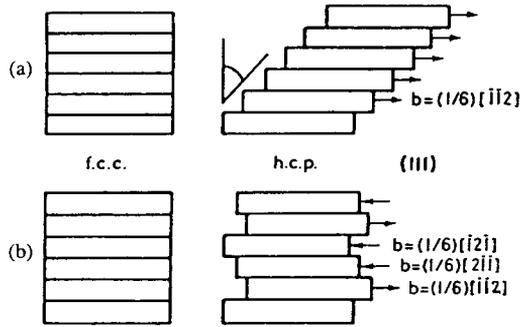


FIG. 21.4. Diagram illustrating two ways of transforming f.c.c. to h.c.p. stacking: (a) a dislocation (or step) with a fixed Burgers vector moves through every alternate plane; (b) dislocations with each of the three possible Burgers vectors are randomly selected to move through alternate planes with the result that there is no macroscopic shape change.

$\{111\}$  plane in a  $\langle 11\bar{2} \rangle$  direction, this being one-half of the twinning shear. As the h.c.p. structure has two atoms in its primitive unit cell, some shuffling would normally be required, but in this case it is accomplished automatically by the atomic mechanism of transformation. This involves simply passing a Shockley partial dislocation through every second  $\{111\}$  plane of the cubic structure, thus converting the  $\dots ABCABC \dots$  stacking to an  $\dots ABABAB \dots$  stacking (Christian, 1951). In the cobalt transformation there is a slight contraction normal to the close-packed planes, which means that the transformation dislocation (see pp. 290–291) is not exactly a Shockley partial but has a component of Burgers vector normal to  $\{111\}$ . A special feature of this transformation is the existence of three equivalent  $\langle 11\bar{2} \rangle$  directions in a  $\{111\}$  plane, the net effect on the atomic arrangement of a displacement in any of these directions being the same. This means that if the successive displacements are not in the same direction but are randomly distributed in all three directions, there will be no macroscopic shape change and, for plates constrained by a surrounding matrix, no associated strain energy (Fig. 21.4); see Fig. 16.23 for the analogous situation in which growth is by diffusion steps.

Experimental work on cobalt and cobalt-nickel alloys shows that fine parallel bands appear on a free surface during spontaneous transformation and detailed analysis shows that each band corresponds to a shape deformation with a displacement in one of the three equivalent  $\langle 11\bar{2} \rangle$  directions of a  $\{111\}$  plane. In cobalt-nickel alloys (Fig. 21.5), the width of each band is about  $1 \mu\text{m}$  at about 25% nickel, decreasing slightly with increasing nickel and more strongly with less than 5% nickel. In pure cobalt, the width of one of these regions of homogeneous distortion may be as little as 10–30 nm, but much wider bands are sometimes found, especially in transformed single crystals (see Fig. 21.5). The individual bands compensate one another on a mesoscopic scale of a few microns and Fig. 21.6 shows a surface profile obtained by atom force microscopy which illustrates this self-compensating effect.

High resolution electron microscopy (Waitz and Karnthaler, 1996) has shown that the bands widen by the motion of elementary transformation dislocations of Burgers vector

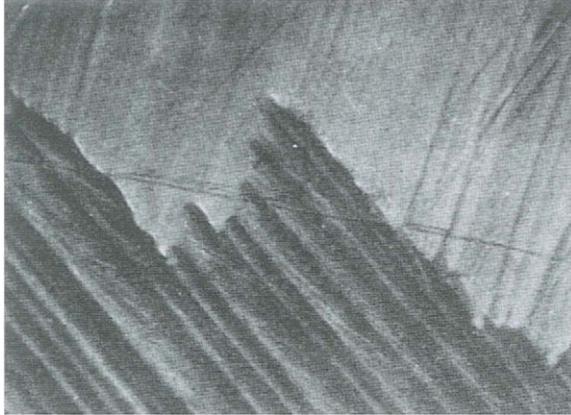


FIG. 21.5. Surface relief at the boundary of two hexagonal orientations in a transformed single crystal of a cobalt-nickel alloy (after Gaunt and Christian, 1959). The bands within each orientation have three distinct tilts corresponding to the three  $\langle 11\bar{2} \rangle$  shear directions.

approximately equal to that of a Shockley partial. The origin of the transformation dislocation steps is uncertain. The wide regions of one shear direction observed in cobalt-nickel could be due to some form of pole mechanism or to some nucleating source, e.g. at a grain boundary where the internal stress favours one particular orientation of the displacement. A possible pole mechanism for the transformation f.c.c.  $\rightarrow$  h.c.p. is shown in Fig. 21.7.

Somewhat different results have been obtained by Hitzenberger and Karnthaler (1991) for the thermally induced reverse transformation on heating where all three Burgers vectors were active and compensation was obtained on an atomic scale. The same result was obtained for transformation in both directions in cobalt-iron alloys where the transformation is f.c.c.  $\rightarrow$  d.h.c.p. structure with ... *ABAC* ... stacking. Waitz and Karnthaler (1997) found that the transformation took place by the motion of superledges or pure steps 12 planes high and with little or no net Burgers vector. These superledges thus produced no overall shear displacement but rearranged the atoms within the step height, so the transformation might alternatively be described as a pure shuffle mode.

Transformation under an externally applied stress will tend to produce one favoured orientation of the Burgers vector, the resulting shape deformation acting as a deformation mechanism. Gaunt and Christian found that when a cobalt-nickel alloy was transformed under stress, the habit plane was not necessarily the  $\{111\}$  plane which had the largest shear stress resolved in a  $\langle 11\bar{2} \rangle$  direction, and this may mean that some kind of pre-existing nucleus larger than a single stacking fault is necessary to initiate transformation. However, the shear direction for transformation under stress always seems to be that one of the three  $\langle 11\bar{2} \rangle$  directions in which the resolved shear stress is greatest. A lenticular plate may be modelled as a series of closed transformation dislocation loops on successive planes, and edgewise growth of such a plate merely involves expansion of these loops. The thickening

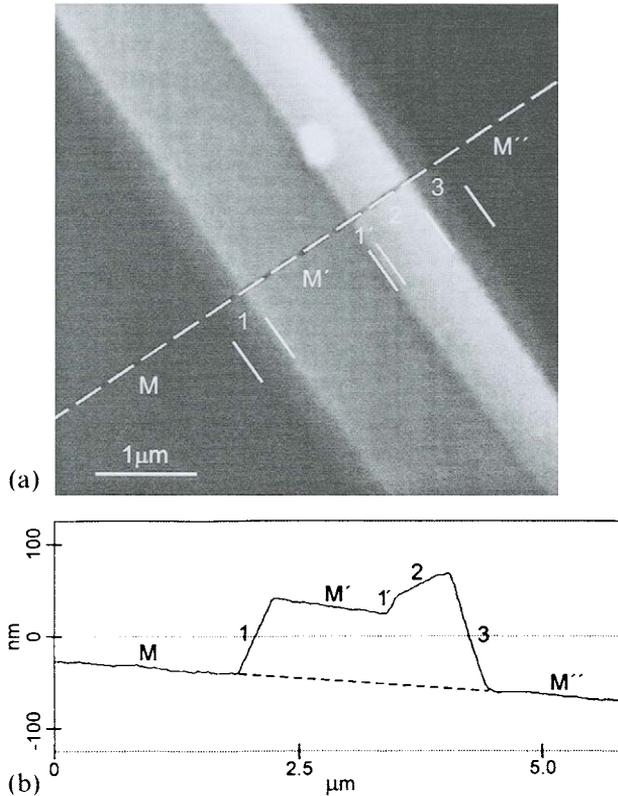


FIG. 21.6. Surface profile showing how three bands mutually compensate their surface distortions (after Waitz and Karnthaler, 1997).

of the plate requires the production of new loops, either spontaneously or by nucleation at favoured sites such as the intersection of the plate with a grain boundary or by some form of pole or cross-slip mechanism. Whatever is the detailed mechanism, there must be mistakes or breaks in the correct sequence of dislocation movements because X-ray and electron microscope observations show that h.c.p. cobalt is very heavily faulted. Spontaneously transformed specimens tend to contain growth faults, whereas specimens transformed under stress contain predominantly deformation faults (Anantharaman and Christian, 1956; Houska *et al.*, 1960).

When the lattice deformation differs from the shape deformation, as in most martensites, it is not possible to give such a detailed description of the atomic movements during transformation. As previously mentioned, the basic requirement of the martensitic interface is that it shall be glissile, and if this is achieved by matching the two structures at the interface with an array of interfacial dislocations, the growth of a plate normal to its habit plane may be regarded as the gliding motion of this dislocation array. This is quite a different model from that just described for fully coherent martensite. The dislocations are

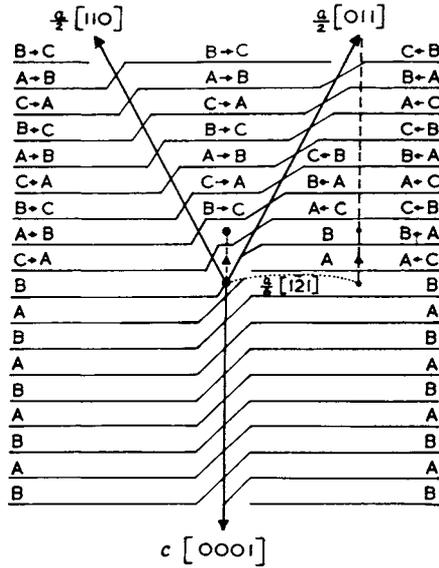


FIG. 21.7. Possible pole mechanism for the transformation f.c.c. → h.c.p.

located in a planar array and do not represent steps in the interface. The Burgers vectors of these dislocations are corresponding lattice vectors when defined in either lattice, whereas that of a transformation dislocation is not a lattice vector. The dislocations will tend to form closed loops normal to the habit plane, in contrast to the loops of transformation dislocations which are parallel to the habit plane. Growth in the normal direction will then involve expansion of the dislocation loops.

A lenticular martensitic plate must contain steps as well as interface dislocations and the edgewise growth of the plate will then take place by the local stress developed at the edge of an extending plate, although no direct evidence is available at present. Growth parallel to the habit plane will presumably be easier in the direction in which existing loops simply have to expand than in the direction in which new loops must be formed, so that the plate will be elongated in one direction rather than being circular.

Experimental evidence is now available that most martensites consist of stacks of fine twins rather than being described by the dislocation model just outlined. The twin intersections with the habit plane then replace the interface dislocations of the above description. This does not affect the conclusions on growth, the only difference being that twins must be nucleated as the plate expands in its own plane.

### 90. DEFORMATION PRODUCED BY MOVING INTERFACES

The glissile motion of a martensitic interface deforms the volume of material through which it sweeps, in much the same way as the motion of a dislocation line. Whereas the



is a second zone axis of undistorted plane normals. The second undistorted plane has unit normal  $\bar{\mathbf{v}}$  where

$$\bar{\mathbf{v}} = (4 + 4g\mathbf{v}'\mathbf{e} + g^2)^{-1/2}(2\mathbf{e} + g\mathbf{v}') \tag{90.4}$$

and the zone axis of the undistorted plane normals is the unit vector

$$\bar{\mathbf{e}} = (4 - 4f\mathbf{v}'\mathbf{e} + f^2)^{-1/2}(2\mathbf{v} - f\mathbf{e}) \tag{90.5}$$

The vectors  $\mathbf{v}$ ,  $\bar{\mathbf{v}}$ ,  $\mathbf{e}$  and  $\bar{\mathbf{e}}$  all lie in the plane  $S$ , as shown in Fig. 21.8;  $S$  is often called the plane of shear. A rotation  $\mathbf{R}$  about the normal to  $S$  through an angle  $\theta$  will simultaneously restore  $\bar{\mathbf{v}}$  and  $\bar{\mathbf{e}}$  to their original orientations. In Fig. 21.8 the (equal) angles between  $\mathbf{v}$  and  $\bar{\mathbf{v}}$  and between  $\mathbf{e}$  and  $\bar{\mathbf{e}}$  are labelled  $\phi$  and the (equal) angles between  $\mathbf{v}$  and  $\bar{\mathbf{e}}$  and  $\bar{\mathbf{v}}$  and  $\mathbf{e}$  are labelled  $(\frac{1}{2}\pi - \psi)$ , and it then follows that

$$\cos \theta = \cos 2\phi = 1 - \frac{1}{2}g^2 \cos^2(\phi + \psi) \tag{90.6}$$

which defines  $\theta$  and  $\phi$  in terms of the original parameters  $\psi$  and  $g$ . The deformation  $\bar{\mathbf{E}} = \mathbf{R}\mathbf{E}$  represents an invariant plane strain of the same magnitude as  $\mathbf{E}$  on the plane  $\bar{\mathbf{v}}$  in the direction  $\bar{\mathbf{e}}$  and the roles of  $\mathbf{v}$  and  $\bar{\mathbf{v}}$  and  $\mathbf{e}$  and  $\bar{\mathbf{e}}$  are interchanged; it may be called the conjugate invariant plane strain. The elements of  $\bar{\mathbf{R}}$  are given from  $\bar{\mathbf{R}} = \bar{\mathbf{E}}\mathbf{E}^{-1}$  as

$$\bar{\mathbf{R}} = \bar{\mathbf{I}} + g\bar{\mathbf{e}}\bar{\mathbf{v}}' - f\mathbf{e}\mathbf{v}' - fg(\bar{\mathbf{v}}'\mathbf{e})\bar{\mathbf{e}}\mathbf{v}' \tag{90.7}$$

These relations are simplified when  $\mathbf{E}$  is a simple shear so that

$$f = g = s \tag{90.8}$$

The angle  $\psi$  is zero for a simple shear and

$$\cos \theta = (4 - s^2)(4 + s^2)^{-1}; \quad \tan \phi = \frac{1}{2}s \tag{90.9}$$

which are familiar results in the theory of deformation twinning.

The elongation or contraction produced by  $\mathbf{E}$  in any direction may be derived as follows. The unit vector  $\mathbf{x}$  is transformed into  $\mathbf{x} + (s\mathbf{d} + \zeta\mathbf{v})\mathbf{v}'\mathbf{x} = \mathbf{x} + \sin\chi_0(s\mathbf{d} + \zeta\mathbf{v})$ , where  $\chi_0$  is the initial angle between  $\mathbf{x}$  and the invariant plane. As the square of the initial length of  $\mathbf{x}$  is given by  $\mathbf{x}'\mathbf{x} = 1$ , the square of the new length is  $1 + 2s\sin\chi_0 \times \cos\lambda_0 + 2\zeta\sin^2\chi_0 + (s^2 + \zeta^2)\sin^2\chi_0$ , where  $\cos\lambda_0 = \mathbf{d}'\mathbf{x}$ , so that  $\lambda_0$  is the initial angle between  $\mathbf{x}$  and  $\mathbf{d}$ . The ratio of final and initial lengths in the direction initially specified by  $\chi_0$  and  $\lambda_0$  is thus

$$l/l_0 = \{1 + 2s\sin\chi_0 \cos\lambda_0 + (s^2 + 2\zeta + \zeta^2)\sin^2\chi_0\}^{1/2} \tag{90.10}$$

and the strain is  $\varepsilon = \Delta l/l_0$ . A simpler form of this equation is obtained if the angle  $\lambda'_0$  between the initial vector direction  $\mathbf{x}$  and the direction  $\mathbf{e}$  is introduced. The change in length becomes

$$l/l_0 = \{1 + 2g\sin\chi_0 \cos\lambda'_0 + g^2\sin^2\chi_0\}^{1/2} \tag{90.11}$$

Equation (90.10) or (90.11) defines the fractional change in length of a tensile specimen which is completely converted from one orientation or structure to another by interface migration. It is a generalization of the well-known equation given by Schmid and Boas

(1936) for simple shear (slip or twinning) and corrects a slightly inaccurate equation given by Otsuka *et al.* (1976) and by Otsuka and Wayman (1977), who simply added  $\xi \sin \chi_0$  to the Schmid and Boas expression. It has been assumed that the ends of the specimen are maintained in alignment; this will usually involve bending of the ends of the specimen which must be excluded from the measurements.

In many experiments, only part of the specimen experiences the shape deformation  $E$ . If, for example, an alternating stack of two structures is converted into a single crystal by consuming one orientation of volume fraction  $z_1$ , eqn. (90.10) gives the change in length if  $s$  and  $\xi$  are replaced by  $z_1 s$  and  $z_1 \xi$ . At any stage in the deformation, when the volume fraction of the whole specimen swept by the interfaces is  $z$ , the change in length is obtained by substituting  $s' = zs$ ,  $\xi' = z\xi$  and  $g' = zg$  for  $s$ ,  $\xi$  and  $g$  respectively in eqns. (90.10) and (90.11).

From eqn. (90.10) or (90.11), it follows that the specimen will increase in length provided that

$$\cos \lambda_0 / \sin \chi_0 > -(s^2 + 2\xi + \xi^2)/2s \quad (90.12)$$

or, equivalently,

$$\cos \lambda'_0 / \sin \chi_0 > -\frac{1}{2}g \quad (90.13)$$

and comparison with eqn. (90.4) shows that this requires  $\mathbf{x}$  to lie within the region bounded by the planes normal to  $\mathbf{v}$  and to  $\bar{\mathbf{v}}$  and on the positive side of both planes (or equivalently on the negative side of both planes). Thus, the projection of  $\mathbf{x}$  on  $S$  lies within the obtuse angle  $QOR$  (or  $POT$ ) in Fig. 21.8. [This angle is obtuse only if the right-hand side of eqn. (90.12) or (90.13) is negative.] The remaining directions, with projections within the angle  $POR$  (or  $QOT$ ), decrease in length. However, these results apply only to specimens after the full shape deformation of eqn. (90.1). During deformation, the macroscopic values of  $s'$  and  $\xi'$  averaged over the whole specimen length gradually increase from zero as  $z$  increases, and some directions within the range of eqn. (90.12) initially decrease in length and then increase again. The condition for continuous increase in length is

$$\cos \lambda'_0 / \sin \chi_0 > 0 \quad \text{or} \quad \cos \lambda_0 / \sin \chi_0 > -\xi/s \quad (90.14)$$

For most thermoelastic martensites,  $|\xi| \ll |s|$  and eqn. (90.14) requires that if  $s$  is positive,  $\lambda_0$  and  $\chi_0$  must both lie between 0 and  $\pi/2$  (or between  $\pi$  and  $3\pi/2$ ), so that the projected direction of the axis must lie in the right-hand upper (or left-hand lower) quadrant of Fig. 21.8 between  $\mathbf{d}$  and  $\mathbf{v}$ . This condition is familiar from the theory of shear deformation by slip.

If both  $\xi$  and  $s^2$  may be neglected in comparison with  $s$ , eqn. (90.10) may be expanded to give the approximate expression

$$\varepsilon = (l - l_0)/l_0 \approx s \sin \chi_0 \cos \lambda_0 \quad (90.15)$$

so that if there are various possible interfaces of differing but equivalent orientations and shape deformation tensors (i.e., crystallographic variants), the one which maximizes

the change in length in the direction of an applied tensile stress will be that for which the “Schmid factor” is greatest. This means that if the normal component  $\xi$  is negligible, the maximum work will be done by the stress field when the resolved shear stress is maximized.

It is sometimes useful to introduce the angles  $\gamma$ , between the specimen axis and the plane  $S$ , and  $\beta$ , between the direction  $\mathbf{d}$  and the projection of the axis on  $S$ . Equation (90.15) then becomes

$$\varepsilon = \frac{1}{2}s \sin 2\beta \cos^2 \gamma \tag{90.16}$$

Alternatively, the angle  $\gamma_0$  between the plane containing  $\mathbf{x}$  and  $\mathbf{d}$  and  $S$  gives

$$\varepsilon = \frac{1}{2}s \sin 2\lambda_0 \cos \gamma_0 \tag{90.17}$$

As a result of the strain, the specimen axis rotates relative to the crystal axes, so that its crystallographic orientation changes. In a manner analogous to that used for slip deformation this change may be shown to be given by

$$\sin \chi / \{(1 + \xi) \sin \chi_0\} = \sin \lambda' \sin \lambda'_0 = l_0 / l \tag{90.18}$$

or, in terms of  $\lambda_0$  and  $\chi_0$ ,

$$\cos \lambda / (\cos \lambda_0 + s \sin \chi_0) = l_0 / l \tag{90.19}$$

If  $\xi$  may be neglected, these equations reduce to the familiar results for slip

$$\sin \chi / \sin \chi_0 = \sin \lambda / \sin \lambda_0 = l_0 / l \tag{90.20}$$

Different expressions are needed for the strain and change of orientation in a compression test on a machine with anvils which maintain the end surfaces of the specimen in a fixed orientation. Thus, if a unit normal  $\mathbf{m}$  is parallel to the stress axis, the ratio of the final and initial lengths of the specimen is  $\{\mathbf{m}'\mathbf{E}^{-1}(\mathbf{E}^{-1})'\mathbf{m}\}^{-1/2}$ . Using eqn. (90.2), this gives

$$l_0 / l = \{1 - 2f \sin \chi_0 \cos \lambda'_0 + f^2 \cos^2 \lambda'_0\}^{1/2} \tag{90.21}$$

which is the generalization of the Schmid and Boas (1936) expression for slip (Bowen and Christian, 1965). The specimen will decrease in length provided

$$\sin \chi_0 / \cos \lambda'_0 < \frac{1}{2}f \tag{90.22}$$

which requires  $\mathbf{m}$  to lie within the region bounded by the planes normal to  $\mathbf{e}$  and to  $\bar{\mathbf{e}}$  and on the positive side of one plane and the negative of the other. The projection of  $\mathbf{m}$  thus lies within  $P'OR'$  or  $Q'OT'$  of Fig. 21.8. The corresponding condition for continuous decrease in specimen length is

$$\sin \chi_0 / \cos \lambda'_0 < 0 \tag{90.23}$$

and for  $|\xi| \ll |s|$  this requires the normal  $\mathbf{m}$  to lie in the left-hand upper or right-hand lower quadrant of Fig. 21.8.

The approximate equations (90.15)–(90.17) also apply to the strain in a compression test but, in crystals of low symmetry, twinning or stress-induced transformation under both

uniaxial tension or compression may not be possible. If the lattice has relatively high symmetry, however, both conditions (90.14) and (90.23) may be satisfied with different, crystallographically equivalent, variants of  $\mathbf{E}$ . A special case arises in cyclic phenomena such as superelasticity and shape memory effects when one structural change  $\alpha \rightarrow \beta$  is reversed, so that the inverse deformation is produced by  $\beta \rightarrow \alpha$ . The matrices  $\mathbf{E}$  and  $\mathbf{E}^{-1}$  of eqns. (90.1) and (90.2) are then interchanged for the reverse stressing, and reverse motion of the original set of interfaces will exactly annul the shape deformation of the forward change.

The rotation of the specimen axis relative to the crystal axes in a specimen compressed between parallel anvils is given by

$$\cos \chi / \cos \chi_0 = (1 + \xi) \cos \lambda' / \cos \lambda'_0 = l/l_0 \quad (90.24)$$

which reduces to the Schmid and Boas expression

$$\cos \chi / \cos \chi_0 = \cos \lambda / \cos \lambda_0 = l/l_0 \quad (90.25)$$

when the deformation is a pure shear.

In contrast to ideal extension or contraction, some elastic strain is inevitable in a bent single crystal specimen, although the bending may sometimes be accomplished without any far-reaching internal stresses. In general, a long rod or bar of length  $l_0$  may be bent in any plane containing its axis to a radius of curvature  $\rho$ , where the maximum curvature attainable by interface displacements is given by

$$1/\rho = \Delta l/(tl_0) = \varepsilon/t \quad (90.26)$$

and  $t$  is the diameter or thickness in the plane of bending. A section of the specimen in the plane of bending is shown in Fig. 21.9(b); the most extended and contracted filaments consist entirely of the orientation or structure favoured by tension and compression respectively. If the specimen originally contains alternating regions with planar boundaries, the deformation is accomplished by rotation of these boundaries out of the ideal invariant planes.

On an atomic scale, the tapering must be accomplished by the introduction of steps into the invariant plane interfaces, i.e. by twinning or transformation dislocations. When the plane of bending is  $S$  of Fig. 21.8 the dislocations are pure edges normal to  $\mathbf{e}$ , and the specimen deforms by uniform plane bending. The top and bottom surfaces of a rectangular section bar then form two concentric cylinders and there is no length change normal to  $S$ . For this limiting case,  $\gamma = \gamma_0 = 0$  and  $\chi_0 = \beta$  so that from eqns. (90.16) and (90.26)

$$1/\rho = (s/2t) \sin 2\beta \quad (90.27)$$

Uniform plane bending is only possible when the plane of the applied bending moment is  $S$ ; in other cases, the specimen appears to twist in addition to bending. The geometry of such a deformation closely parallels the deformation attainable by single glide of dislocations along a set of initially parallel slip planes.

Consider a set of twins or martensite plates which taper continuously in the direction  $\mathbf{d}$  to end in lines perpendicular to  $S$  (Fig. 21.10).  $\mathbf{E}$  will now be approximated as a pure shear; the strain fields are more complex if  $\xi$  is appreciable. If the angle of taper is small, the interfaces consist of coherent sections with widely spaced edge type twinning or

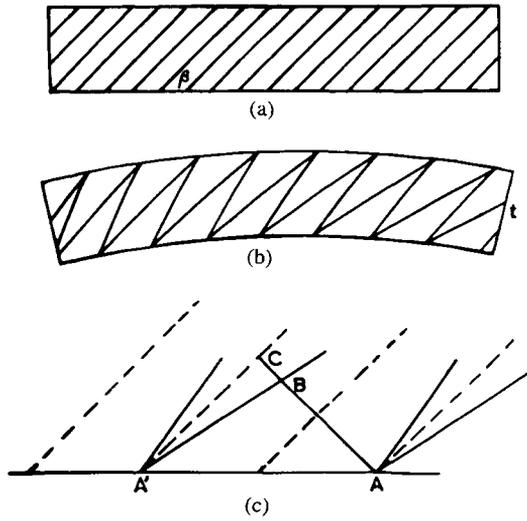


FIG. 21.9. Bending of single crystal specimen by rotation of glissile interfaces: (a) trace of invariant planes in unbent crystal; (b) rotated invariant planes in bent specimen; (c) two adjacent plates (or twins) meeting in compressed surface.

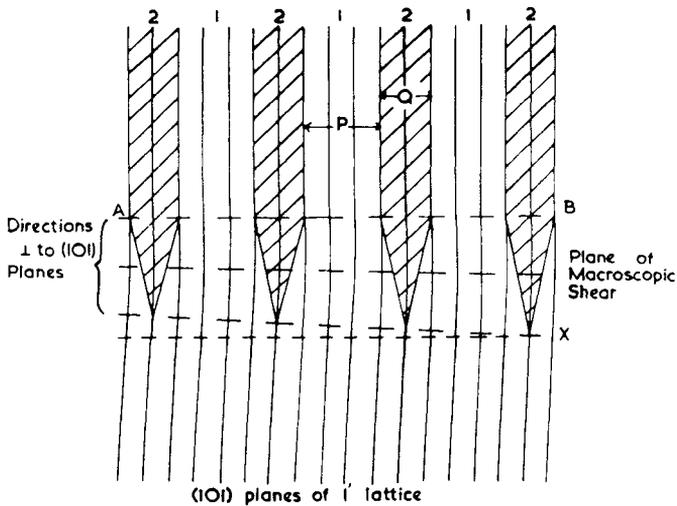


FIG. 21.10. Tapering twins producing distributed shear. The gradual elimination of orientation 2 produces orientation 1' which is slightly rotated from orientation 1. The particular crystallography shown refers to  $\{101\}$  tetragonal twins.

transformation dislocations. The lattice planes parallel to the interface thus become curved in the tapered region and, because the Burgers vector of the twinning dislocation lies in the invariant plane, the distance between any two lattice planes of this type must remain constant if accumulating long-range stresses are to be avoided.

Let the original mean thicknesses of the two orientations or structures be  $p$  and  $q$  and consider two neighbouring sections normal to the interfaces where the mean thicknesses of structure 1 are  $p_1$  and  $p_2$  respectively. If there are  $2n$  interfaces, the total number of step dislocations between the two sections is  $n(p_2 - p_1)/h$ . Each step produces a rotation  $b/(n(p + q))$ , and the total rotation between the two sections is thus

$$\begin{aligned}\alpha_1 - \alpha_2 &= b(p_2 - p_1)/(h(p + q)) \\ &= s(p_2 - p_1)/(p + q)\end{aligned}\quad (90.28)$$

As a result of the rotation, the duplex region 1 + 2 changes into the reorientated lattice  $1'$  which is rotated about the normal to  $S$  through an angle

$$\alpha = qs/(p + q)\quad (90.29)$$

with respect to lattice 1.

This result may now be applied to the bending of a single crystal bar shown schematically in Fig. 21.9. Suppose the bar has rectangular cross-section and one face in the plane  $S$  of the applied bending moment so that tapering is produced by edge type steps. Each lattice will have constant orientation along a line such as  $ABC$  of Fig. 21.9(c), so there will be a change of orientation between  $A$  and  $A'$  which from eqn. (90.28) will be  $s(p + q)\cos \beta/t$ , where  $p$  and  $q$  were the original thicknesses of the two regions. The mean distance between  $A$  and  $A'$  is  $(p + q)/\sin \beta$ , so that the total misorientation between the ends of the bottom surface is  $(ls/2t)\sin 2\beta$ , which corresponds to the radius of curvature given by eqn. (90.27). The neutral plane, defined as the plane containing equal proportions of the two regions after bending, is the central plane of the deformed specimen, and the thickness changes during the bending unless  $p = q$  initially. The lattice rotation relative to any filament of length  $l'$  is given by

$$\sin \beta' / \sin \beta = l/l' = \rho_c/(\rho_c + y)\quad (90.30)$$

where  $\rho_c$  is the radius of curvature of the neutral filament and  $y$  is measured in the thickness direction from an origin in the neutral filament. This equation was first derived by Nye (1953) for the analogous problem of a bar bent by an excess density of lattice edge dislocations on one set of slip planes, and it implies that the lattice planes parallel to the invariant plane are the involutes of a circle.

Now consider the bending of a rectangular bar in which the top surface again contains the normal to  $S$  but the side surface is inclined to  $S$ . If the interfaces taper by edge type dislocations, the specimen again lies between the surfaces of two coaxial cylinders of radii  $\rho_c \pm 0.5t$ , but the specimen axis is now inclined to the cylinder axis, and the top surface of a long specimen appears to twist as well as bend. In the plane of the applied bending moment, the projected radius of curvature is

$$\rho = \rho_c^2 \sec^2 \gamma\quad (90.31)$$

at the centre of the specimen and, for a sufficiently short specimen [ $l \sin 2\gamma \ll \rho_c$ , i.e.  $1/t \ll (s \sin 2\gamma \sin 2\beta)^{-1}$ ], the section will approximate to a circle with this curvature. Thus, to a first approximation, directions along the length of the specimen are still bent into circular arcs of radius  $\rho = tl/\Delta l$ .

### 91. EFFECTS OF STRESS; THERMOELASTIC TRANSFORMATION

As discussed briefly in Section 3, a martensitic transformation is usually athermal, a given fraction of the total volume becoming martensite on cooling to a particular temperature below the  $M_s$  temperature at a rate which is independent of temperature. This is the behaviour normally observed in steels, and each plate then forms in a very short period of time. On subsequent cooling, new plates are formed but there is no general increase in the size of existing plates. The amount of transformation is thus a function of temperature, and is related to the number of nuclei which are activated in any given temperature interval and to the size to which they grow.

Isothermal martensite formation was first reported by Kurdjumov and Maximova (1948) for a steel containing 9% manganese and 0.6% carbon, and later (Kurdjumov and Maximova, 1950) for an iron-nickel-manganese alloy. They found that transformation could be completely suppressed by rapid cooling to  $-180^\circ\text{C}$  and that isothermal transformation then occurred on reheating to temperatures in the range  $-80$  to  $-100^\circ\text{C}$ . The transformation kinetics are thus very different from those of normal athermal martensite and C-curve behaviour is observed on a  $T-T$  diagram. These results were repeated by Cech and Hollomon (1953) using a nickel-manganese steel of slightly different composition. Almost completely isothermal transformation was obtained in the temperature range  $-70$  to  $-100^\circ\text{C}$ , as shown in Fig. 21.11. Subsequently, Shih *et al.* (1955) obtained almost completely isothermal transformation at  $-196^\circ\text{C}$  and higher temperatures, and they suggested that small amounts of martensite formed athermally may have a considerable effect on the form of the isothermal transformation curves.

Kurdjumov concluded that nucleation rather than growth was the rate-determining step in isothermal martensite formation and metallographic observations confirmed that individual plates form very rapidly as in athermal martensite. Conclusive evidence that the difference in kinetic behaviour is not attributable to the growth mechanism was provided by the work of Bunshah and Mehl (1953) on the velocity of formation of individual plates in an iron-nickel-carbon alloy. Using an electrical resistivity method in conjunction with equipment with a very fast response, they were able to show that individual plates form in about  $10^{-7}$  s and that the linear growth velocity is about  $1 \text{ km s}^{-1}$  at all temperatures in the range  $-20$  to  $-200^\circ\text{C}$  for both athermal and isothermal martensite. This growth velocity is of the order of magnitude of the velocity of sound, as also are observed speeds of fast crack propagation. The high velocity at low temperatures and the lack of temperature dependence prove that the growth is not thermally activated.

Many transformations are known in which athermal martensite formation is followed by some isothermal reaction. In most cases, it is probable that the isothermal reaction

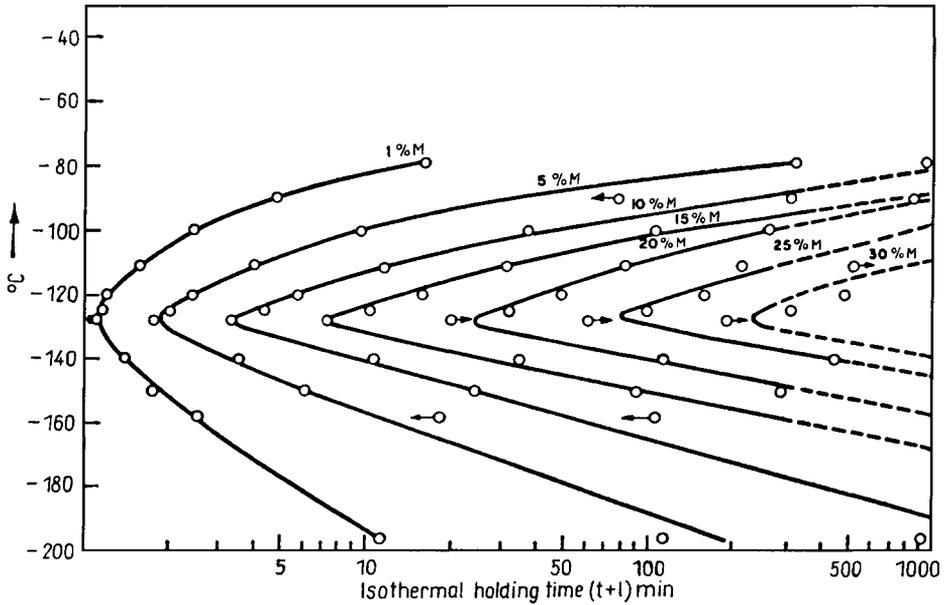


FIG. 21.11. Isothermal transformation curves for martensite formation in an iron-nickel-manganese alloy (after Cech and Hollomon, 1953).

involves the nucleation of new plates but Yeo (1962, 1964) demonstrated that there can be true isothermal growth of martensite in nearly carbon-free iron-nickel alloys. Presumably some relaxation process at or ahead of the interface controls the growth rate. In general, the overall kinetics of martensite formation are not much influenced by the growth mechanism because of the very rapid growth rate. They are, however, influenced by the nucleation rate because an individual plate does not continue to grow indefinitely but reaches some maximum size. In distinguishing martensitic transformations from other reactions, the growth mechanism is of primary importance and reliance on kinetic characteristics is unsatisfactory precisely because of their sensitivity to nucleation rates. In athermal martensite formation, the operational nucleation rate is a function of temperature but not of time, whilst in isothermal nucleation the number of nuclei increases with time at a particular transformation temperature.

There is considerable experimental evidence to show that martensite nuclei do not form randomly in the parent phase but only at preferred sites. In many transformations which are reversible on heating, the original grain structure is regained after completion of a cycle of transformation. If the specimen is then cooled again to give a second transformation, the position of the product regions and the order in which they form is repeated almost exactly and this complete reversibility of structure can be repeated through very many cycles of transformation. Such observations are particularly striking in transformations like that in manganese-copper alloys where the transformed regions have parallel sides and retain their mobilities. The observation may be taken to mean that

preferred sites of fixed potency are present in the parent phase or that the product regions of the first transformations are not completely eliminated on heating, so that ready-made nuclei are available for all subsequent transformations. In any case, the reversibility can be destroyed by annealing at a high temperature well above the normal transformation range. Rather similar effects are found in single crystals of gold–cadmium or indium–thallium alloys which can be induced to transform by a single interface from only one nucleus after prolonged heating near the melting point. On subsequent transformation cycles, transformation always begins from the same place and with the same kind of interface (there are several possibilities). However, a further annealing treatment will change the mode of transformation.

Cech and Turnbull (1956) used the successful small droplet technique to obtain evidence in support of the theory of heterogeneous nucleation. They austenitized small particles of iron–nickel alloys and quenched them into the martensitic transformation range. They found wide variations in the  $M_s$  temperatures of individual particles and some did not transform at all. Sufficiently large particles ( $> 100 \mu\text{m}$ ) all had the same  $M_s$  temperature as the bulk alloy of the same composition. These results are interpreted on the assumption that some particles contained suitable nucleating agents of varying potency whereas others did not.

When a finite region of matrix transforms into a martensite plate, the changes of shape and volume required by the transformation mechanism have to be accommodated by elastic or plastic deformation in the surrounding matrix. The existence of this accompanying deformation may complicate the measurement of the shape change of the transformation, but in principle it is quite distinct from the shape change itself. As a plate grows within a matrix, the change of shape will lead at first to purely elastic deformation, but clearly the stresses may reach the elastic limit if the plate grows to a sufficient size, and plastic deformation will then ensue.

This deformation may take one of the forms shown schematically in Fig. 20.15 and may be responsible for the cessation of growth. The plastic deformation will normally be non-reversible and will remain after the plate is retransformed. If it destroys the coherency of the interface, further growth will be impossible, even under an increased driving force. It seems probable that this is why martensite plates once formed often do not grow further nor shrink on reversal of the driving force. Instead a fresh nucleus of the parent phase must form within the transformed plate.

As in the case of twins, it is also possible for a plate to stop growing before the elastic limit is reached, and this leads to the phenomenon known as thermoelastic transformation. There is now essentially an equilibrium between the mechanical and chemical forces, which wish to expand the plate, and the coherency stresses, which are resisting further expansion. Thermoelastic transformations were first observed by Greninger and Mooradian (1938) and studied in detail by Kurdjumov and Khandros (1949). Figure 21.12 illustrates the very different behaviour on cycling through the transformation on cooling and heating of a thermoelastic alloy (gold–cadmium) and a non-thermoelastic alloy (iron–nickel). The thermoelastic hysteresis is very much smaller than that of the iron–nickel alloy.

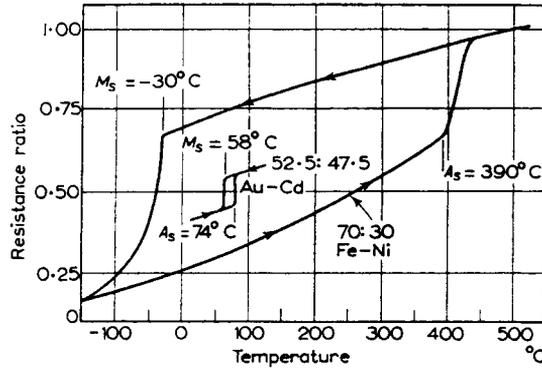


FIG. 21.12. Comparison of cyclic transformations in thermoelastic alloy (Au-Cd) and non-thermoelastic alloy (Fe-Ni) (after Kaufman and Cohen, 1958).

The theory of thermoelastic equilibrium is approximately as follows. Neglecting the surface energy which is only important at the nucleation stage, the change in free energy when a martensite plate forms consists of a negative term proportional to the volume of the plate and a positive term, representing the stored elastic energy. If the shape stays constant, the elastic energy term is also proportional to the volume, as is usually assumed in nucleation theory. Eventually, however, growth in the edgewise direction may cease because of grain boundaries and other obstacles and further increase in volume then changes the shape and causes the strain energy to rise more rapidly than the chemical energy decreases. The total free energy may then pass through a minimum value as the plate grows and, if this happens before the elastic limit is reached, the plate is in equilibrium with the matrix. An increase in the chemical driving force, e.g. by cooling, will then displace the equilibrium to a larger size of plate, and a decrease will reduce the volume of the plate. The plate will thus grow or shrink as the temperature is changed, and similar effects could be produced by variation of an externally applied stress.

The above treatment of thermoelastic equilibrium considers only the equilibrium of an isolated plate and Olson and Cohen (1975) showed that the equilibrium condition for a particle of dimensions  $R$ ,  $R$  and  $y$  ( $\partial G/\partial y = 0$ ) corresponds to  $\Delta g^{\text{AM}} + 2A(y/R) = 0$ , where  $\Delta g^{\text{AM}}$  is the chemical driving force and  $2A(y/R)$  is twice the stored elastic energy. However, a more complete treatment must include the possibility of nucleation as well as plate growth due to a change of temperature and stress (Tong and Wayman, 1975). A typical thermoelastic loop differs from a non-thermoelastic transformation both in the extent of the hysteresis and in the relative locations of  $A_s$  and  $M_s$  the start temperatures for the transformation on heating and cooling respectively. In the non-thermoelastic case,  $M_s < A_s$ , and a long-standing estimate (Kaufman and Cohen, 1958) of  $T_0$ , the temperature at which the two phases have equal chemical free energies, is  $\frac{1}{2}(M_s + A_s)$ . Both the shear and hydrostatic components of an applied stress field may interact with a martensitic plate and verification of the effects of a hydrostatic pressure alone was first made by Kulin *et al.* (1952) for iron-nickel alloys. A quantitative prediction of the effects

of a given stress may be made if it is assumed that the  $M_s$  temperature is characterized by a critical value of the driving force, and calculations of this type are in good agreement with experiment. The mechanical work done by an applied stress field  $\sigma$  per unit volume of the product phase is

$$\Delta g_{\text{mech}} = \sigma_n \xi + \tau s \quad (91.1)$$

where  $\sigma_n = \mathbf{n}' \boldsymbol{\sigma} \mathbf{n}$  and  $\tau = \mathbf{d}' \boldsymbol{\sigma} \mathbf{n}$  are the interface stress components in the directions  $\mathbf{n}$  and  $\mathbf{d}$  respectively. The work done by the external stress may be added to the chemical free energy change to give a net free energy change for the transformation. The first term may have either sign but the generally much larger second term will normally be positive in crystals of high symmetry for some variants of the habit plane. For an applied uniaxial tensile or compressive stress,  $\sigma_a$ , if  $\chi_0$  is the initial angle between the specimen axis and the invariant plane and  $\lambda_0$  is the initial angle between the specimen axis and the direction of displacement  $\mathbf{e}$  then, at the beginning of transformation,

$$\Delta g_{\text{mech}} = \sigma_a (\xi \sin^2 \chi_0 + s \sin \chi_0 \cos \lambda_0) \quad (91.2)$$

Equations (91.1) and (92.2) are essentially equivalent to Schmid's law modified slightly to allow for a deformation mechanism which gives an invariant plane strain instead of a simple shear. The net driving force

$$\Delta g = \Delta g_{\text{chem}} + \Delta g_{\text{mech}} \quad (91.3)$$

may now be assumed to have a critical value at the temperature  $M_s^\sigma$  at which transformation begins. As  $\Delta g_{\text{chem}}$  is approximately linear in temperature, the temperature  $M_s^\sigma$  will rise approximately linearly with stress. Patel and Cohen (1953) found quantitative agreement between theory and experiment in iron-nickel alloys.

Shear stresses, unlike pressures, are not true state variables, so their use in thermodynamic equations is subject to purist objections. However, if solid-state viscous flow is forbidden, thermodynamic parameters may be defined with the same justification as for any metastable state. In practice, this means that irreversible processes such as atomic diffusion are negligible on the timescale of the experiments.

A second reservation is that the effects of stress may sometimes be independent of those of temperature. For example, in a stress-free transformation, variants often form self-accommodating groups to reduce the strain energy whereas, under stress, a single preferred variant will dominate. Thus the stored energy which may consume much of the driving force will be different in the two cases. A more extreme example is that some transformations under stress are not observed in the absence of stress. This may readily be understood if a potential second martensite phase exists which is metastable at all temperatures in the absence of stress, but which has a larger shape deformation than the stable martensite. The larger value of  $\Delta g_{\text{mech}}$  may then ensure that at a sufficient stress the second martensite phase forms either directly from the parent or from the first martensite.

Some authors (see, e.g., Otsuka *et al.*, 1976; Schroeder and Wayman, 1978; Otsuka and Shimizu, 1979) have produced tentative phase diagrams with temperature and shear stress

as independent variables for those alloys in which new phases appear under applied shear stresses.

A third reservation is that part of the driving force at  $M_s$  may be required for the nucleation of the new phase so that an implicit assumption is that  $\Delta g_{\text{mech}}$  contributes to nucleation as well as to growth. If the critical nucleus configuration differs from that of a fully formed plate, this may be incorrect. Some authors have cited results on the effects of stress on the transformation in single crystals of an iron–nickel–carbon alloy as evidence that the interaction of the stress with the lattice deformation rather than the shape deformation is the determining factor. The correct relationship is, however, not very readily deduced from the experimental results, and may vary in different alloy systems.

The linear relation between  $M_s^\sigma$  and  $\tau$  predicted above is analogous to the Clausius–Clapeyron equation (Burkart and Read, 1953) and, as the work done by the applied stress per unit volume of the new phase is also approximately  $\sigma_A \varepsilon$ , the equation may be written

$$d\sigma_s/dT_o = N\Delta h/T_o\varepsilon = N\Delta s/\varepsilon \quad (91.4)$$

where  $N$  is the number of atoms per unit volume, and  $\Delta h$  and  $\Delta s$  are the enthalpy and entropy per atom respectively. The enthalpy includes the mechanical work and is given by

$$\Delta h = \Delta g + T\Delta s = \Delta h_{\text{chem}} - \sigma_a \varepsilon \quad (91.5)$$

The assumption that the  $M_s^\sigma - T_o^\sigma$  relation is linear then gives

$$dM_s^\sigma/d\sigma_a = \varepsilon/\Delta s \quad (91.6)$$

Instead of the free energy equation, it may be more convenient to represent the chemical driving force as an equivalent stress, so that

$$\Delta g_{\text{chem}} = \sigma_{n,\text{chem}}\xi + \tau_{\text{chem}}s \cong \tau_{\text{chem}}s \quad (91.7)$$

(see, e.g., Patel and Cohen, 1953). In crystals of low symmetry, some applied stresses may oppose the transformation for all possible variants of the habit plane. The variant chosen is then that for which this effect is least; the effective driving force is reduced and transformation does not begin until the temperature is reduced below the normal  $M_s$  temperature.

The effects just mentioned arise from a change in the equilibrium temperature between the austenite and martensite phases in the presence of stress, and are normally small in the range of laboratory stresses. In some alloys, much larger changes in  $M_s$  are obtained by applying stresses beyond the elastic limit so that some deformation of the specimen takes place. When a specimen is stressed in this way, transformation may occur at still higher temperatures, the stress required increasing as the temperature is raised. This is often referred to as strain-induced transformation (as opposed to stress-aided transformation, which is the thermodynamic effect discussed above). Eventually an upper limit  $M_d$  is reached above which the specimen cannot be transformed. If nucleation and growth reactions to equilibrium phases do not intervene, the reverse transformation on heating may also be martensitic and similar effects are then observed, the transformation beginning at temperatures below  $A_s$ . The raising of  $M_s$  to  $M_d$  and the lowering of  $A_s$  to  $A_d$  may be regarded as a kinetic effect which enables the forward and reverse reactions to take

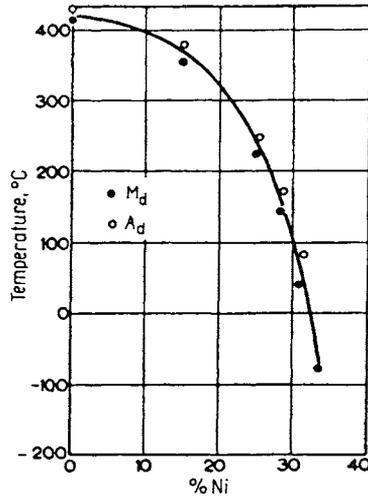


FIG. 21.13. Stress-induced transformations on cooling and heating in cobalt nickel alloys (after Hess and Barrett, 1952).

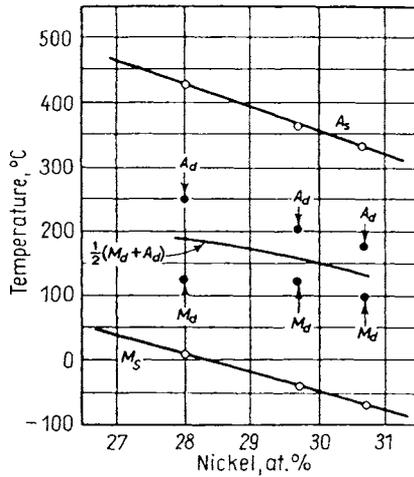


FIG. 21.14. Transformation temperature for spontaneous and stress-induced transformations on cooling and heating in iron-nickel alloys (after Kaufman and Cohen, 1956).

place at lower driving forces. It then seems reasonable to bracket the  $T_0$  temperature between the two, i.e. at  $\frac{1}{2}(M_d + A_d)$ , and in some cases the two temperatures are almost coincident. Figure 21.13 shows results of Hess and Barrett (1952) for cobalt-nickel alloys where this condition is fulfilled. More often there remains some appreciable difference between  $M_d$  and  $A_d$ . Figure 21.14 shows some results of Kaufman and Cohen (1956) for

iron–nickel alloys in which the 400°C gap between  $A_s$  and  $M_s$  is reduced to 100°C between  $A_d$  and  $M_d$ .

In a typical thermoelastic cycle of cooling and heating, the inverse transformation on heating begins at a lower temperature than  $M_s$  because of the remaining strain energy which assists the reverse change on heating. Clearly, the approximation  $T_0 = \frac{1}{2}(M_s + A_s)$  cannot then be valid as it would place  $T_0$  below  $M_s$ , which is thermodynamically impossible. Instead Tong and Wayman suggested that  $T_0 = \frac{1}{2}(M_s + A_f)$ , which places  $T_0$  below  $A_f$ . However, Olson and Cohen concluded from an analysis of the thermodynamic equilibrium of an isolated plate that a position of instability would arise on shrinkage at a negative value of the net driving force, so that the last plate would disappear completely at  $A_f$  which is necessarily below  $M_s$ . Wayman (1982) has pointed out that the difference in the two conclusions arises from the different assumptions about the stored elastic energy at the beginning of the cooling transformation. The first plate to form on cooling will be the last plate to remain on heating. Olson and Cohen effectively assumed that the plate in the form of an oblate spheroid would have appreciable elastic energy whereas Tong and Wayman assumed that the elastic energy close to  $M_s$  will be negligible.

The macroscopic deformation of a specimen in which glissile interfaces are displaced depends upon the extent to which the ideal shape deformation associated with each interface is cancelled or “accommodated” by the deformation produced by another interface and on the extent to which it is constrained by a surrounding matrix. At one extreme, a planar interface traverses the whole cross-section of a single crystal, dividing it into two regions, each of which is either a single crystal or finely twinned. Displacement of such an interface generates no stress field and its mobility will be limited only by a frictional force analogous to the Peierls–Nabarro force on a dislocation. In the other limit, a plate may grow entirely within the matrix, and the opposing elastic stresses may then cause growth to cease at some finite driving force or may induce some irreversible plastic deformation.

The martensitic transformation of a single crystal by the movement of a single planar interface from one end of the specimen was first observed in gold–cadmium alloys by Chang and Read (1951) and similar observations were made shortly afterwards for indium–thallium alloys (Burkart and Read, 1953; Basinski and Christian, 1954) (see Fig. 21.15). Single interface transformations are usually obtained by heating or cooling the specimen in a temperature gradient so that the interface is always close to the equilibrium transformation temperature, and the experiment is akin to the Bridgman technique for growing single crystals from the melt. The velocity of the interface is thus controlled mainly by the heating or cooling rate, although it is often jerky on a fine scale, presumably because of the presence of localized obstacles. The product (low temperature phase) in both cases is twinned on a fine scale, the twins being visible in an optical microscope except in the immediate vicinity of the interface. Many attempts have been made to obtain single interface transformations in other materials but these were all unsuccessful until Otsuka *et al.* (1973) succeeded in transforming a copper–aluminium–nickel alloy in this way. Single interfaces may also be stress-induced and measurement of  $\tau_{cif}$  then gives the frictional stress on the moving pair of interfaces at a particular velocity which is governed

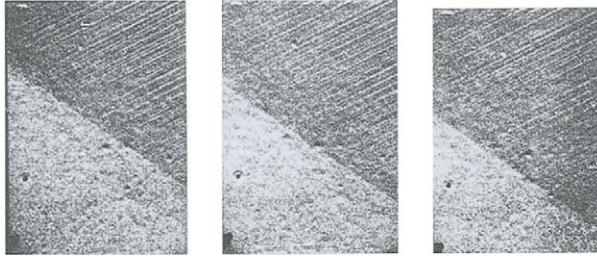


FIG. 21.15. Single interface transformation in indium–thallium alloy. Successive frames from a movie film of the interface are shown.

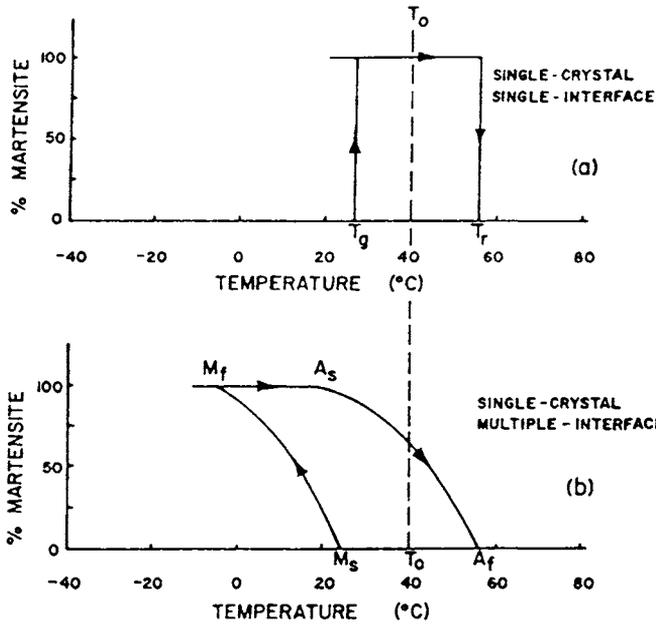


FIG. 21.16. Transformation hysteresis in a single crystal of Cu 14% Al 2.5% Ni (after Salzbrenner and Cohen, 1979): (a) single interface transformation; (b) multiple interfaces.

by the applied strain rate. The corresponding measurement of the thermal hysteresis in a stress-free transformation may be combined with calorimetric data to give  $\Delta g_{chem}$  and hence  $\tau_{eff}$ . Figure 21.16(a) shows the results obtained by Salzbrenner and Cohen (1979) for a copper–aluminium–nickel alloy;  $M_s = M_f$  and  $A_s = A_f$ , and the small temperature hysteresis represents twice the frictional stress on the interface. The deduced value of the friction stress is stated to agree reasonably well with the value for a comparable interface velocity measured by stress hysteresis.

Salzbrenner and Cohen confirmed that in a single crystal of copper–aluminium–nickel transforming by multiple interfaces (obtained by cooling uniformly rather than in a

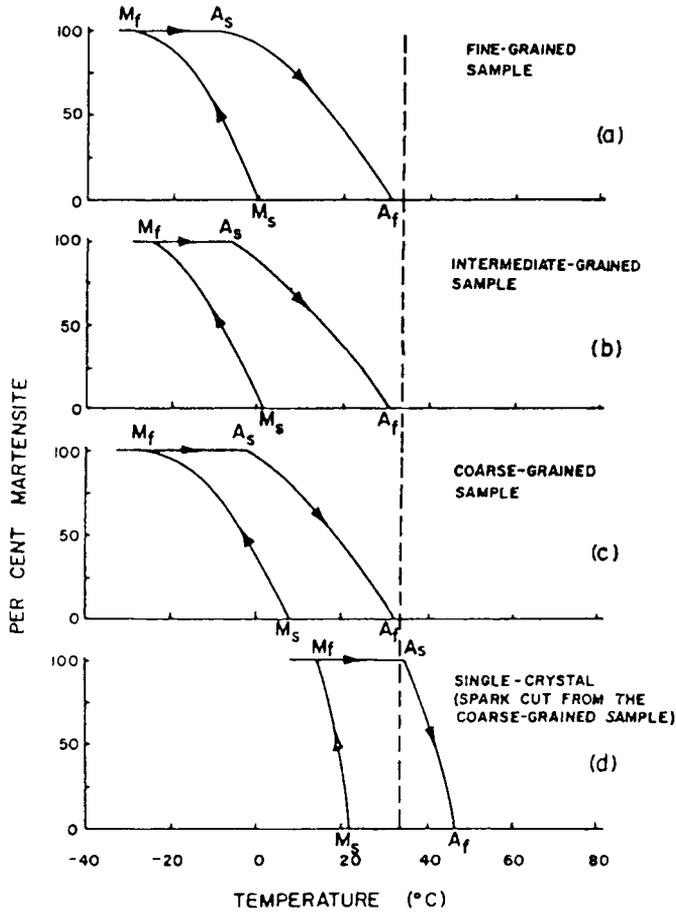


FIG. 21.17. Transformation hysteresis for polycrystalline Cu 14% Al 2.5% Ni (after Salzbrener and Cohen, 1979). Grain sizes: (a) 0.5 mm; (b) 1.5 mm; (c) 4 mm; (d) cut out single crystal.

temperature gradient) the  $M_s$  and  $A_f$  temperatures are almost the same as those for single interface transformation, but the  $M_f$  and  $A_s$  temperatures are significantly lower [see Fig. 21.16(b)] as later plates to form do so in this case with appreciable strain energy. Transformation hysteresis curves for polycrystalline specimens of the same alloy (Fig. 21.17) show that the transformation range extends to lower temperatures with decreasing grain size because of the increasing elastic constraints, and in all cases  $A_f$  is equal to or slightly below  $T_0$ , as the friction stress is small. The thermoelastic behaviour is, however, almost removed again in a single crystal cut out of the coarse-grained sample, and this experiment illustrates the considerable influence of grain boundary constraints on interface transformation and deformation properties.

The compensation of the shape change of one plate by the simultaneous or subsequent formation of other variants is called self-accommodation. All stress-free transformations tend to be self-accommodating because the shape change of any plate will produce stress fields favouring the formation of other variants, whereas transformation under stress will tend to produce plates of a single variant. In non-thermoelastic alloys, the internal stresses round a fully transformed plate, although partly relaxed by plastic deformation, will nevertheless tend to promote nucleation of other variants, an effect known as "autocatalysis" or "sympathetic nucleation". An extreme example is the effect known as the "burst phenomenon" in steels. At a temperature a little way below  $M_s$ , there is a very rapid transformation of an appreciable fraction of the whole volume of the specimen, and metallographic evidence shows that the plates, much thicker than in normal transformations, have formed a rough zigzag pattern in the specimen. The process is somewhat analogous to a chain reaction, each plate in turn producing a stress field to activate the nucleus of the next plate. The whole group forms a partly self-accommodating system and this is presumably why individual plates are thicker than normal. Machlin and Cohen (1951) found that about 25.5% of the whole transformation in an iron-nickel alloy was produced in a burst of very short duration, accompanied by an audible click.

Many (possibly all) thermoelastic alloys transform to give plates in rather efficient self-accommodating groups which allow virtually complete transformation with almost zero average shape change. The simplest form of self-accommodation is in alloys such as manganese copper and indium-thallium where adjacent parallel-sided plates or "bands" have opposite shear directions and cancel each other. Another special case, already discussed, is the f.c.c.  $\rightarrow$  h.c.p., d.h.c.p., etc. transformations in which cancellation of the shear component occurs in groups of three and sometimes on an atomic scale.

Self-accommodation in many copper alloys is achieved by sets of four variants which form a diamond-shaped or a parallelogram-shaped pattern grouped round a  $\{110\}$  plane of the parent structure which makes a comparatively small angle with the habit planes (see Chapter 26). The exact arrangement depends on the detailed crystallography, but the net result in many cases is that virtually total transformation can be obtained without significant overall deformation and the various transformed regions are separated from each other by habit plane or intervariant interfaces which are still glissile. Such a configuration leads to shape memory effects (see Section 105).

In addition to single interface transformation, indium-thallium single crystals undergo a mode of transformation termed the  $X$  interface. This consists of two single interfaces on  $\{110\}$  type planes at  $60^\circ$  which cross each other. Each interface with the parent phase leads to a twinned product, with one orientation (c) common to the two product regions. The  $\{110\}$  habit plane of each set of (c) twins is parallel to the other parent-twin interface and the continuation of the two parent-product interfaces into the product region gives two interfaces between twinned and untwinned products, the final configuration being a single crystal of product. The interfaces  $XOB$  and  $AOO'$  in Fig. 21.18 represent an  $X$  interface and  $OO'Y$  and  $COZ$  represent another type of crossing interface (an  $X'$  interface) with a common portion  $OO'$ . Each of the twinned regions in the figure contains two tetragonal

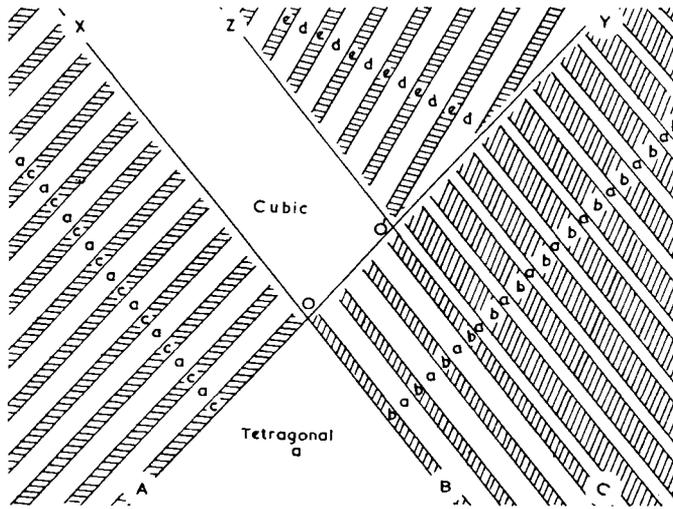


FIG. 21.18.  $X$  and  $X'$  crossing interfaces in indium thallium alloy (after Basinski and Christian, 1954).

orientations in the volume ratio 2:1; these orientations and their relative proportions are indicated by small letters and shading respectively. The shears along  $XOB$  and  $ZO'C$  are necessarily in the same sense. The orientations (a), (c), (d) and (e) represent four of the eight possible orientations of twins.

Yet another type of crossing interface, the  $\lambda$  interface, was observed by Basinski and Christian. Three of the four half interfaces are sharp and correspond to changes in the twinning pattern as discussed above. However, the fourth half interface is diffuse and is a tapering interface of the type shown in Fig. 21.19. When one of the twin orientations has finally been eliminated, the lattice rotation of the other is just sufficient to make its orientation identical to that of the single crystal obtained from the  $X$  part of the interface. The disposition of twins is indicated in Fig. 21.19.

If the cooling or heating of a single interface is interrupted for a time and then resumed, the interface does not resume its motion immediately but waits until a finite driving force has been established. This is the single interface analogue of the well-known effect of thermal stabilization in polycrystalline specimens. Thermal stabilization is just such an experiment; the specimen is held for a period of time at a temperature below  $M_s$  and when cooling is resumed it is found that the transformation does not begin at once. When transformation resumes, it often does so at an enhanced rate and some martensite may form in bursts (Glover and Smith, 1956). Thermal stabilization may affect athermal and isothermal martensite, but it is not found in steels unless carbon or nitrogen is present. The question of whether or not thermal stabilization occurs on holding above the  $M_s$  temperature is rather controversial; widely differing results have been recorded.

Stabilization is normally measured as the temperature interval needed to cause a resumption of transformation; this is sensitive both to the holding time and the

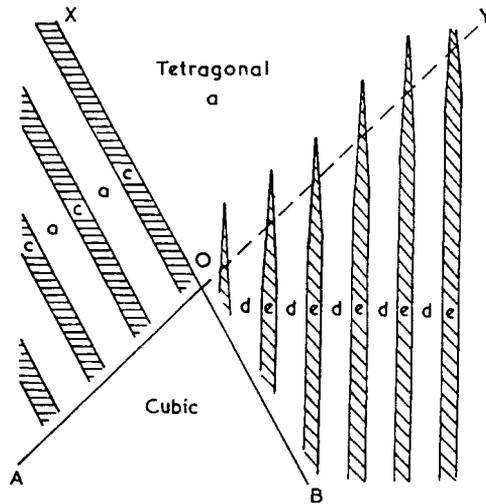


FIG. 21.19.  $\lambda$  interface in indium thallium alloy (after Basinski and Christian, 1954).

temperature. The behaviour is very complex and varies with the composition of the alloy. Work by Edmondson (1957) and by Glover (1962) shows that the maximum stabilization obtainable after long holding times is greater for holding temperatures near the bottom of the transformation range than it is for temperatures high in the range. In contrast, the rate of transformation after stabilization is very rapid for low temperature holding, so that the curve of volume fraction transformed versus temperature rapidly approaches that for unstabilized specimens. After a high temperature stabilizing treatment, the effect is more permanent.

Theories of stabilization attribute the effect either to changes in the untransformed austenite or to changes in the way in which already transformed regions affect the still untransformed austenite. In the first category are theories in which the matrix dislocations are pinned by precipitates or solute atmospheres, thus hardening the structure and making direct growth and accompanying plastic deformation more difficult. The time-temperature characteristics of stabilization are qualitatively compatible with this kind of theory, originally suggested by Morgan and Ko (1953), but experimental activation energies (Philibert, 1955; Woodilla *et al.*, 1959) are very much smaller than the energy for diffusion of interstitials in austenite. This suggests that, in the alloys investigated, the operative process may be carbon or nitrogen diffusion in the martensite, the activation energy of  $15.5 \pm 3$  kcal g atom<sup>-1</sup> obtained by Woodilla *et al.* being reasonably consistent with this hypothesis. Obviously both types of effect may contribute to the overall result and there may be many different phenomena to be considered under this general heading.

There are two kinds of stabilization which require the presence of some volume of martensite. One of these, favoured by Woodilla *et al.*, is that interstitial atoms migrate towards the interfaces between austenite and martensite and lock them in position. This is quite analogous to the locking of a single interface. The other possibility is that the

autocatalytic effect of existing martensite is destroyed or reduced during stabilization. Glover and Smith (1956) suggested that interstitial atoms could so distribute themselves as to reduce these elastic fields which are thought to be responsible for the autocatalysis. Glover (1962) suggested that this is the main effect in all normally observed types of stabilization. Early objections to this theory were based mainly on activation energies although, as pointed out above, the energies agree much better than those which suppose that diffusion in austenite is the rate-determining step. In any event, the more detailed results of some later investigations show that the kinetics of the whole process are very complex and it is doubtful whether they could be described as having a simple rate-determining step with a fixed activation energy.

If stabilization depends upon removing the internal stresses around previously formed plates, the presumed transformation should have some similar characteristics to the initial transformation at  $M_s$ . Glover pointed out that alloys which show an initial burst just below  $M_s$  behave similarly after stabilization. Also, Philibert and Crussard (1955) showed that, after stabilization, an isothermal transformation in a high carbon steel proceeded at a very slow rate characteristic of a fully isothermal transformation, rather than at the much faster rate normally observed when isothermal martensite is preceded by some athermal martensite. As remarked on p. 977, the athermal martensite apparently has a considerable effect on the subsequent isothermal change.

In addition to thermal stabilization, it is possible to stabilize mechanically. Application of stress above the  $M_d$  temperature results only in plastic deformation.

#### REFERENCES

- ANANTHARAMAN, T. R. and CHRISTIAN, J. W. (1956) *Acta Crystallogr.* **9**, 479.  
 BASINSKI, Z. S. and CHRISTIAN, J. W. (1954) *Acta Metall.* **2**, 101, 148.  
 BIBRING, H., SEBILLEAU, F., and BUCKLE, C. (1958-9) *J. Inst. Met.* **87**, 71.  
 BILBY, B. A. and CHRISTIAN, J. W. (1956) *The Mechanism of Phase Transformations in Metals*, p. 121, The Institute of Metals, London; (1961) *J. Iron Steel Inst.* **197**, 122.  
 BOWEN, D. K. and CHRISTIAN, J. W. (1965) *Philos. Mag.* **12**, 369.  
 BOWLES, J. S. and MACKENZIE, J. K. (1954) *Acta Metall.* **2**, 129.  
 BUNSHAH, R. F. and MEHL, R. F. (1953) *Trans. Metall. Soc. AIME* **197**, 251.  
 BURKART, M. W. and READ, T. A. (1953) *Trans. Metall. Soc. AIME* **197**, 315.  
 CECH, A. E. and HOLLOMON, J. H. (1953) *Trans. Metall. Soc. AIME* **197**, 685.  
 CECH, A. E. and TURNBULL, D. (1956) *Trans. Metall. Soc. AIME* **206**, 124.  
 CHRISTIAN, J. W. (1951) *Proc. R. Soc. London, Ser. A* **206**, 51.  
 EDMONDSON, B. (1957) *Acta Metall.* **5**, 208.  
 GAUNT, P. and CHRISTIAN, J. W. (1959a) *Acta Metall.* **7**, 529; (1959b) *Ibid.* **7**, 534.  
 GLOVER, S. G. (1962) *J. Iron Steel Inst.* **200**, 102.  
 GLOVER, S. G. and SMITH, T. B. (1956) *The Mechanism of Phase Transformations in Metals*, p. 265, The Institute of Metals, London.  
 GRENINGER, A. B. and MOORADIAN, V. G. (1938) *Trans. Metall. Soc. AIME* **128**, 337.  
 HAYZELDEN, C., HATTOPADHYAY, K., BARRY, J. C., and CANTOR, B. (1991) *Philos. Mag.* **A63**, 461.  
 HESS, J. B. and BARRIETT, C. S. (1952) *Trans. Metall. Soc. AIME* **194**, 645.  
 HOUSKA, C. R., AVERBACH, B. L., and COHEN, M. (1960) *Acta Metall.* **8**, 81.  
 HULL, D. (1954) *Bull. Inst. Met.* **2**, 134.  
 KAUFMAN, L. and COHEN, M. (1956) *Trans. Metall. Soc. AIME* **197**, 1251; (1958) *Prog. Met. Phys.* **7**, 165.

- KO, T. and COTTRELL, S. A. (1952) *J. Iron Steel Inst.* **172**, 307.
- KULIN, S. A., COHEN, M., and AVERBACH, B. L. (1953) *Trans. Metall. Soc. AIME* **194**, 661.
- KURDJUMOV, G. V. (1948) *J. Tech. Phys. SSSR* **18**, 999.
- KURDJUMOV, G. V. and KHANDROS, L. C. (1949) *Dokl. Akad. Nauk SSSR* **66**, 211.
- KURDJUMOV, G. V. and MAXIMOVA, O. P. (1948) *Dokl. Akad. Nauk. SSSR* **61**, 83; (1950) *Ibid.* **73**, 85.
- MACHLIN, E. S. and COHEN, M. (1951) *Trans. Metall. Soc. AIME* **191**, 746.
- MORGAN, E. R. and KO, T. (1953) *Acta Metall.* **1**, 36.
- NYE, J. F. (1953) *Acta Metall.* **1**, 153.
- OLSON, G. B. and COHEN, M. (1975) *Scr. Metall.* **9**, 247.
- OTSUKA, K., TAKAHASHI, M., and SHIMIZU, K. (1973) *Metall. Trans.* **4**, 2003.
- OTSUKA, K. and WAYMAN, C. M. (1977) *Reviews of the Mechanical Behaviour of Materials*, Vol. II, ed. P. Feltham, p. 81, Freund, Tel Aviv.
- OTSUKA, K., WAYMAN, C. M., NAKAI, K., SAKAMOTO, H., and SHIMIZU, L. (1976) *Acta Metall.* **24**, 207.
- OTTE, H. M. and READ, T. A. (1957) *Trans. Metall. Soc. AIME* **209**, 412.
- PATEL, J. R. and COHEN, M. (1953) *Acta Metall.* **1**, 531.
- PHILIBERT, J. (1955) *C.R. Acad. Sci.* **240**, 190.
- PHILIBERT, J. and CRUSSARD, C. (1955) *J. Iron Steel Inst.* **180**, 39.
- SALZBRENNER, R. J. and COHEN, M. (1979) *Acta Metall.* **27**, 739.
- SCHMID, E. and BOAZS, W. (1936) *Kristallplastizität*, Springer-Verlag, Berlin.
- SCHROEDER, T. A. and WAYMAN, C. M. (1978) *Acta Metall.* **26**, 1746.
- SHIH, C. A., AVERBACH, B. L., and COHEN, M. (1955) *Trans. Metall. Soc. AIME* **203**, 183.
- TAKEUCHI, S. and HONMA, T. (1957) *Sci. Rep. Res. Inst., Tohoku Univ., Ser. A* **9**, 492, 508.
- TONG, H. C. and WAYMAN, C. M. (1975) *Acta Metall.* **23**, 209.
- TROIANO, A. R. and GRENINGER, A. B. (1946) *Met. Prog.* **50**, 303.
- WAITZ, T. and KARNTHALER, H. P. (1996) *Philos. Mag.* **73A**, 365; (1997) *Acta Mater.* **45**, 837.
- WAYMAN, C. M. (1982) *Solid-Solid Phase Transformations*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 1119, Metallurgical Society of AIME, Warrendale, PA.
- WOODILLA, A. J., WINCHELL, P. G., and COHEN, M. (1959) *Trans. Metall. Soc. AIME* **215**, 849.
- YEO, R. B. G. (1962) *Trans. Metall. Soc. AIME* **224**, 1222; (1964) *Trans. Am. Soc. Met.* **57**, 48.

## CHAPTER 22

# *Crystallography of Martensitic Transformations*

### 92. GENERAL DESCRIPTION OF FORMAL THEORIES

Formal theories of martensite crystallography make predictions about the crystallographic features of a particular transformation from a knowledge of the two crystal structures and their coexisting lattice parameters together with an assumption about the plane and/or direction of the lattice-invariant deformation. The various theories differ slightly in their assumptions and in the mathematical representations made, but they are all essentially equivalent and may be described together. Considerable success has been achieved in explaining the rather complex experimental results, although many details remain to be clarified.

Early theories of martensite attempted to show how one structure could be produced from the other by means of a simple shear or other deformation process, combined with local atomic readjustments. These theories were unsuccessful because they ignored the problem of fitting the two structures together and maintaining contact, thus allowing one to grow at the expense of the other. As mentioned on pp. 61–62 Jaswon and Wheeler concluded that the habit plane of a growing martensite plate is an unrotated plane of the deformation relating the two structures. Although it is always possible to find such a plane, vectors within this plane are rotated in general; as already stated all vectors in an invariant plane must be unrotated and unchanged in magnitude during the transformation. A true matching plane must be an invariant plane of the deformation and, as already noted, the condition for that, namely that one principal strain be zero, is very restrictive. The recognition that this condition cannot be fulfilled for any of the infinity of possible deformations carrying one lattice into the other is the basis of the formal theories. These theories postulate that the structures need not fit together on an atomic scale provided that the atomic misfit does not accumulate, so that there is an average fit over all macroscopic dimensions.

The foundation of the theory was developed by Greninger and Troiano (1940, 1949). In their important experimental and theoretical paper, they pointed out that the shape deformation could not effect the change of lattice and suggested that the lattice change should be regarded as the resultant of two successive shears. The first of these shears was then supposed to cause the change of shape, whilst the second shear has no macroscopic effects. This description is incomplete and in particular it does not allow for the volume

change of the transformation, but it is closely similar to the later theories. Greninger and Troiano applied their theory only to their own results on an iron–nickel–carbon steel, but two shear theories of other transformations followed. In some respects this was unfortunate as the shears were supposed to operate consecutively rather than simultaneously, and this concept persists in some of the literature. The difficulty concerning the volume change was tackled by Bowles (1951), who showed that the shape deformation need not be a simple shear but may be any invariant plane strain. All obstacles to the production of a general theory had then been removed.

The first general theory to be published (Wechsler *et al.*, 1953) dealt with transformations in steels but was applied in later papers to gold–cadmium and indium–thallium alloys. A slightly different theory was developed independently by Bowles and Mackenzie (1954) in general form, and subsequently applied to steels and to a wide class of cubic–orthorhombic and cubic–hexagonal transformations. A later development was the surface dislocation theory of Bullough and Bilby (1956) which is formally equivalent to its predecessors, but again uses a rather different form of mathematical analysis. These three theories will be abbreviated to WLR, BM and BB respectively in what follows.

The first stage in all descriptions is to select a lattice correspondence for the transformation, that is, to specify two unit cells which are to be interrelated by the lattice deformation. The correspondence automatically fixes the directions of the principal axes and the magnitudes of the principal strains of the lattice deformation; it specifies what is sometimes called the Bain strain as it was Bain who first proposed the pure lattice strain which relates austenite and martensite in steels.

No general rule for selecting the correspondence is known but the important factors seem to be the size of the smallest unit cells which are related by the lattice deformation and the magnitude of the principal strains. Wherever possible, the two unit cells related by  $S$  will be the smallest consistent with the requirement that each must contain the same number of atoms. If this number is  $n$ , the lattice deformation will leave only a fraction  $1/n$  of the atoms on their correct sites, and the remaining atoms will have to shuffle. If  $n$  becomes large, it is increasingly difficult to give physical significance to  $S$ , even though a larger unit of transformation might allow a reduction in the magnitude of the principal strains.

The transformations which have been analysed theoretically support this conclusion. The correspondence assumed is then that which minimizes the principal strains for the unit of transformation. With one exception, these correspondences have been found by inspection, and the principal axes of the lattice deformation are then parallel to rational crystallographic axes of the parent structure. This implies that the lattice transformation converts an orthorhombic, tetragonal or cubic unit cell of the parent into an orthorhombic, or higher symmetry, unit cell of the product structure. It is usually evident whether or not two such cells of similar size and shape exist in the two structures; in known transformations, the maximum principal strains deduced in this way are about 20%.

There is no theoretical necessity for the principal axes to be rational lattice vectors and, in the case of the  $\beta - \alpha$  transformation in uranium, they are not. In this transformation, the  $\beta$  unit cell contains 30 atoms, so that the homogeneous lattice deformation can

describe only a small fraction of the atom movements. In this case a trial-and-error procedure has to be used to find the correspondence. From the discussion on pp. 58–59, it follows that the columns of the correspondence matrix must be rational lattice vectors of the product structure if the unit of transformation is one primitive cell of the parent, half lattice vectors of the product if the unit of transformation is two unit cells of the parent, and so on. The determinant of the correspondence matrix is equal to the ratio of the number of atoms in the unit cells defined by the bases in both structures. The possible correspondence matrices for any given unit of transformation may thus be written down. There are many possibilities when the primitive cells differ appreciably in size.

An analytical procedure is now needed to define the principal axes and strains for any given correspondence. In Section 7, the procedure for finding these quantities for any representation  $\mathbf{S} = (\mathbf{A}\mathbf{S}\mathbf{A})$  of the strain  $\mathbf{S}$  was given. The correspondence does not define  $\mathbf{S}$  uniquely but it does define the pure deformation associated with  $\mathbf{S}$  and hence a knowledge of  $\mathbf{C}$  is sufficient to diagonalize  $\mathbf{S}'\mathbf{S}$  as on p. 49.

Using  $\mathbf{A}$  and  $\mathbf{M}$  as bases in the parent phase and martensite respectively, then from the definition of the correspondence

$$(\mathbf{A}\mathbf{S}\mathbf{A}) = (\mathbf{A}\mathbf{J}\mathbf{M})(\mathbf{M}\mathbf{C}\mathbf{A})$$

and from eqn. (6.21)

$$(\mathbf{A}^*\mathbf{G}\mathbf{A}) = (\mathbf{A}\mathbf{J}'\mathbf{M})(\mathbf{M}^*\mathbf{G}\mathbf{M})(\mathbf{M}\mathbf{J}\mathbf{A})$$

Substituting these two equations into eqn. (7.23) then gives

$$(\mathbf{A}\mathbf{C}'\mathbf{M})(\mathbf{M}^*\mathbf{G}\mathbf{M})(\mathbf{M}\mathbf{C}\mathbf{A}) - \eta_i^2(\mathbf{A}^*\mathbf{G}\mathbf{A})[\mathbf{A}; \mathbf{u}] = 0 \quad (92.1)$$

as the equation defining the principal vectors. For non-trivial solutions

$$|(\mathbf{A}\mathbf{C}'\mathbf{M})(\mathbf{M}^*\mathbf{G}\mathbf{M})(\mathbf{M}\mathbf{C}\mathbf{A}) - \eta_i^2(\mathbf{A}^*\mathbf{G}\mathbf{A})| = 0 \quad (92.2)$$

This equation can be solved when the correspondence and the metrics of the two phases are known, and the principal vectors and magnitudes of the principal strains are then obtained as described in Section 7.

For any arbitrarily assumed correspondence, it will generally be found that at least one of the principal strains is unrealistically large. A correct correspondence may thus be identified by systematically listing the possibilities and calculating the principal strains for each. Lomer (1956a) used this method to examine possible correspondences in the uranium transformation.<sup>†</sup> He tried 1600 possibilities in which the unit of transformation was a primitive unit cell of the  $\beta$  structure and found only two admissible correspondences in which the principal strains were all less than 20%. Two correspondences were identified in which two  $\beta$  unit cells were the unit of transformation

<sup>†</sup>Lomer used a slightly different approximate procedure from that given above, which does not require any approximations. See Lomer (1956b) and Christian (1956).

but experimental results were used as a guide as an exhaustive search would have been difficult in this case.

When the correspondence has been selected, it only remains to specify how the lattice deformation is converted into the shape deformation. All of the original theories assume that the shape deformation is given by a combination of the lattice deformation and a lattice-invariant simple shear. The simple shear may be a slip shear, produced by an array of dislocations in the habit plane interface, or a twinning shear, in which case the product is a stack of fine twins. As the twinning hypothesis is equivalent to a slip shear on the  $K_1$  plane in the  $\eta_1$  direction, it will not be necessary to consider the twinning case separately. The physical question of what determines whether a dislocation array in the interface or the production of a set of fine twins is the mechanism for the lattice-invariant deformation is discussed later.

The detailed assumptions about the lattice-invariant shear have varied somewhat in the different theories. In the early WLR papers, the product was assumed to be a stack of fine twins or, alternatively and equivalently, a lattice-invariant shear on the twinning plane and in the twinning direction of the product. In later papers, the possibility of using other rational slip planes and directions was explored, although not systematically. Bowles and Mackenzie started from the experimental observation that martensite plates sometimes contain macroscopic twins with the same shape deformation as the matrix. Because the resultant of two invariant plane strains can itself be an invariant plane strain only if the invariant plane or the direction of displacement is common to both, it then follows that the lattice-invariant shear must be a shear either on the observed twinning plane or in the twinning direction. The twins were also assumed to have equivalent correspondence with the matrix, from which it follows that the twinning plane of the product must be derived from a mirror plane (for type I twins) and the  $\eta_1$  direction must be derived from a twofold axis (for type II twins) of the parent. Thus although Bowles and Mackenzie assume that the structures fit together macroscopically without buckling, they impose severe restrictions on the choice of plane and the direction of the shear.

The different procedures used in the various theories will now be outlined. In each case the input data required are the crystal structures and lattice parameters of the two phases, the correspondence and an assumption about the lattice-invariant shear. The equivalence of the different theories may be demonstrated either algebraically or geometrically (Bilby and Christian, 1956; Christian, 1955-6); both descriptions will be used. In general, the algebraic approach is more suitable to calculation but the geometric approach gives the better physical picture. Various graphical methods, involving manipulations on a stereographic net, have also been described by Lieberman *et al.* (1957) and by Lieberman (1969). These procedures are equivalent to the geometrical description but are not discussed further here.

The WLR procedure is to calculate the effect of the pure lattice deformation and the lattice-invariant deformation with an arbitrary amount of shear, and then to impose the condition that the resultant deformation shall contain an undistorted plane.

Now it is required to make the matrix  $\mathbf{F}$  represent a deformation in which there are two planes which are undistorted although rotated. This is true if one of its principal strains is

zero, i.e. one of its principal deformations  $\lambda_i$  is equal to unity. This means that the characteristic equation

$$|\mathbf{F}'\mathbf{F} - \lambda_i^2| = 0 \quad (92.3)$$

must have one root of unity. The components of  $\mathbf{F}$  may all be written in terms of the principal deformations  $\lambda_i$ , the plane and direction of the lattice-invariant shear and the unknown amount of shear  $k^M$ . Substitution of  $\lambda_i^2 = 1$  leads to a quadratic equation in  $k^M$ , giving two different solutions  $k_+^M$  and  $k_-^M$  in the general case. In general, these solutions can be quite different, but when certain symmetry conditions are satisfied they become crystallographically equivalent variants of one solution.

The geometrical significance of the above procedure is illustrated in Fig. 22.1, which shows a unit sphere and the ellipsoid produced from it (the first ellipsoid) by the deformation  $\mathbf{P}$ . The vectors  $\mathbf{l}$  and  $\mathbf{l}^M$  represent the direction of the lattice-invariant shear before and after the deformation  $\mathbf{P}$  and  $\mathbf{m}$  and  $\mathbf{m}^M$  similarly represent the directions of the normals to the plane on which the shear takes place before and after  $\mathbf{P}$ . The first ellipsoid is given a shear of magnitude  $k^M$  on the plane  $\mathbf{m}^M$  in the direction  $\mathbf{l}^M$  to produce a new ellipsoid which is called here the shape ellipsoid. The amount of shear is adjusted so that the shape ellipsoid touches the unit sphere on one of its principal axes.

In order to simplify Fig. 22.1, it has been assumed that a principal axis of the lattice deformation is normal to the plane of the lattice-invariant shear, i.e. is normal to  $\mathbf{m}$  and  $\mathbf{l}$ . Although this is a special assumption, it corresponds to some important real cases. The problem is then two-dimensional (the figure is rather complex for the general three-dimensional case) and the shape ellipsoid and the unit sphere touch along a direction in the plane of shear. The two possible amounts of shear are illustrated in Fig. 22.1(a) and (b).

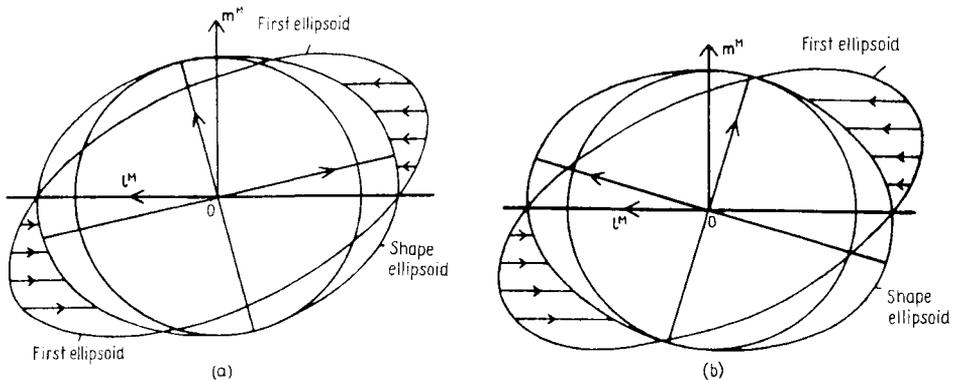


FIG. 22.1. Production of an undistorted plane by shear of the "first" ellipsoid. The images are sections in the plane of shear, which is assumed to be normal to a principal axis of the first ellipsoid.  $\mathbf{m}^M$  and  $\mathbf{l}^M$  are vectors along the normal to the invariant plane of the shear and along the shear direction respectively, both in their positions after the pure deformation. The two possible amounts of shear are shown in (a) and (b).

The principal axes of the shape ellipsoid are equal in length and are related by a 180° rotation about  $\mathbf{m}^M$ . This is not necessarily true in the general case.

The deformation  $\mathbf{F}$  converts the unit sphere into the shape ellipsoid. As this is not a pure deformation, the axes of the shape ellipsoid and the reciprocal shape ellipsoid do not coincide; the latter are the principal axes of the shape deformation. The equation of the reciprocal strain ellipsoid is

$$\mathbf{u}'\mathbf{F}'\mathbf{G}\mathbf{F}\mathbf{u} = 1 \tag{92.4}$$

or

$$G_{ik}F_{il}F_{kj}u_lu_j = 1$$

and its axes are found by solving the discriminating cubic equation which is identical in form to eqn. (7.25). In an orthonormal system, eqn. (92.4) has the form of eqn. (7.19) and the condition for real solutions becomes eqn. (92.3).

The possible habit planes relative to the product lattice are given by the planes of intersection of the shape ellipsoid and the unit sphere. These planes contain the principal axis of unit length and one of the lines of unchanging length in the lattice deformation. Equation (9.45) showed that there is a cone of such lines provided the  $\eta_i$  have not all the same sign. The intersection of the plane  $\mathbf{m}^M$  with the cone provides two lines of unchanging length in the deformation  $\mathbf{F}$ . It follows that a condition for a solution is that the plane  $\mathbf{m}^M$  does intersect the cone.

More conveniently, the habit plane relative to the parent lattice is given by the planes of intersection of the reciprocal shape ellipsoid and the unit sphere. There are two reciprocal shape ellipsoids, corresponding to the two values of the shear  $k^M$ ; in the simple case of Fig. 22.1 these are related by a 180° rotation about  $\mathbf{m}$ . In the orthonormal coordinate system  $\mathbf{J}$  defined by the axes of the reciprocal shape ellipsoid, the habit plane normal  $\mathbf{v}$  has components

$$[\mathbf{J}; \mathbf{v}] = (1 + C^2)^{-1/2} \begin{bmatrix} \pm C \\ 1 \\ 0 \end{bmatrix} \tag{92.5}$$

and the equations of the two habit planes are

$$\bar{j}_1 = \pm C \bar{j}_2 \tag{92.6}$$

(see Fig. 22.2). Both of these planes are undistorted; either may be made invariant by combining the deformation  $\mathbf{F}$  with a suitable rotation  $\mathbf{R}$ . A specified set of principal axes and shear components leads to four different solutions for the habit plane, corresponding to the two values of the lattice-invariant shear and the choice of sign in eqn. (92.6). Each of these habit planes is associated with a particular orientation relation, so that in the general case there are four sets of possible habits and orientation relations, each with up to 24 crystallographically equivalent variants.

In certain cases, the geometry becomes degenerate in the sense that different solutions are equivalent variants of each other. There may then be only two habit planes, or even one plane, corresponding to any particular assumption about the lattice-invariant

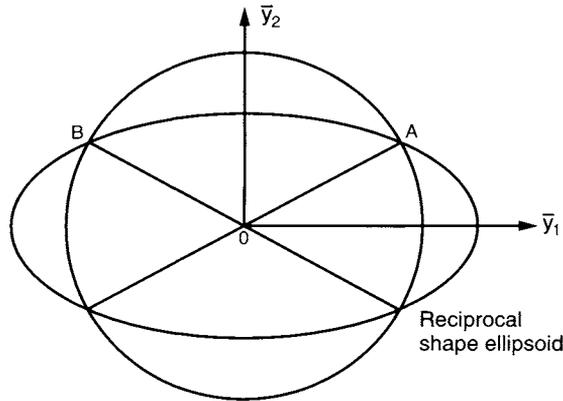


FIG. 22.2. Section through the reciprocal strain ellipsoid. The unit sphere and the reciprocal strain ellipsoid touch along  $\bar{y}_3$ , normal to the figure, so that  $OA$  and  $OB$  are traces of undistorted planes, either of which may be made invariant by adding a suitable rotation.

deformation. This aspect of the theory was examined by Wechsler (1959) and by Christian (1955-6); only a brief description will be given here.

Consider first the choice of sign in eqn. (92.6). In general, the habit planes and orientation relations will be different for the two solutions but they may become equivalent if a principal axis of the reciprocal shape ellipsoid coincides with a symmetry element of the parent crystal. Figure 22.2 shows a section of the reciprocal shape ellipsoid in the  $\bar{y}_1\bar{y}_2$  plane. The traces of the possible habit planes are related to each other by reflections in both the  $\bar{y}_1$  and the  $\bar{y}_2$  lines and, if the two planes are to be equivalent, the two-dimensional point group symmetry must include two mirror lines and hence be at least ( $2mm$ ).

Consideration of the projection symmetries of the three-dimensional point groups then shows that either  $\bar{y}_1^M$  or  $\bar{y}_2^M$  must be a twofold symmetry axis. This conclusion is also evident, as the habit planes may be seen to be related by a  $180^\circ$  rotation about either of these axes. If the planes are to be equivalent, such a rotation must bring the crystal axes into self-coincidence.

In practice, it is clear that, apart from coincidental equivalence depending on the lattice parameters, the conditions will be satisfied only when a principal direction of the lattice deformation is coincident with a twofold axis of the parent phase, and the plane of the lattice-invariant shear is normal to this axis. The two solutions of eqn. (92.6) are then related by a reflection in this plane of shear, and this is the situation in the simple case shown in Fig. 22.1.

The degeneracy which results if the two values of the lattice-invariant deformation give equivalent results is not so easy to discuss. In Fig. 22.1 the principal shape deformations are the same for  $k_+^M$  and  $k_-^M$  and the principal axes are related by a  $180^\circ$  rotation about the plane normal  $\mathbf{m}$ . From this it follows that intrinsic degeneracy results if  $\mathbf{m}$  is a twofold symmetry axis. Clearly, there is no degeneracy possible if the two values of the shear do

not give equivalent shape deformations, as is possible for a general choice of  $\mathbf{m}$  and  $\mathbf{l}$ . However, in some circumstances, degeneracy may result if  $\mathbf{l}$  is a twofold axis and  $\mathbf{m}$  is not, as was shown by Wechsler (1959) in a detailed discussion of this problem for a particular set of principal axes of the lattice deformation. A rather more general discussion is given by Wechsler and Otte (1961).

The twinning hypothesis used by BM and in some of the WLR papers requires that the ( $K_1$ ) invariant plane of the lattice-invariant strain be derived from a mirror plane of the parent structure (for type I twins) or that the ( $\eta_1$ ) displacement direction of the lattice-invariant shear be derived from a twofold axis of the parent.<sup>†</sup> Moreover, when either  $K_1$  or  $\eta_1$  is chosen, the other is selected automatically. This result depends on the assumption that the two solutions have equivalent variants of one lattice correspondence.

Consider two different lattice deformations in which identical principal strains are applied to axes related by a  $180^\circ$  rotation about  $\mathbf{m}$ . These will give the same product only if they represent equivalent variants of one correspondence so that  $\mathbf{m}$  must be a twofold axis of the parent structure. The positions of the plane normals  $\mathbf{m}^M$  after the lattice deformation will not be identical as the rotation from  $\mathbf{m}$  to  $\mathbf{m}^M$  will be equal and opposite in the two cases. However, if the two first ellipsoids are given equal and opposite rotations, so that the two plane normals are made to coincide again, the product lattices which they used to represent will clearly be rotated by  $180^\circ$  about  $\mathbf{m}$ . These two lattices are now twins with the normal to the twin plane given by  $\mathbf{m}^M$ .

Lattice-invariant shears of magnitude  $k_+^M$  and  $k_-^M$  applied after one lattice deformation will give the same reciprocal shape ellipsoids as the lattice-invariant shears  $-k_+^M$ ,  $-k_-^M$  applied after the other lattice deformation. The corresponding pairs of shape ellipsoids differ only by the relative rotation needed to make the two directions of  $\mathbf{m}^M$  coincide. A given shape ellipsoid may thus correspond to either of two twin-related lattice orientations, so that it follows that one habit plane may correspond to either of two twin-related lattices. Moreover the shape deformation is identical for these two lattices so that twins may be present together in a macroscopic plate.

Because a shape ellipsoid is obtained by giving one first ellipsoid a shear  $k_+^M$  or an ellipsoid representing a twinned lattice a deformation  $k_-^M$ , it follows that the two values of the shear are related by

$$k_+^M + k_-^M = s \quad (92.7)$$

where  $s$  is the twinning shear converting one orientation into the other. Care must be taken with this equation when signs are considered; it refers to the possible shears given one first ellipsoid to yield equivalent but different habit planes. When  $0 < k_-^M < k_+^M < s$ , these two shears are in the same direction as in Fig. 22.1. The two shears given to two different first ellipsoids to yield the same habit plane are  $k_+^M$ ,  $-k_-^M$  or  $k_-^M$ ,  $-k_+^M$ ; these are in opposite directions if  $0 < k_+^M < k_-^M < s$ .

<sup>†</sup>A mirror plane is equivalent to a twofold axis normal to the plane only in centrosymmetric structures. The reader is reminded that the theory of the present section applies only to lattices so that the distinction between twofold axes and mirror planes in symmetry operations or twinning laws loses its significance.

Now consider the possibility of having a martensite plate which is a stack of fine twins. If the stack consists of alternate regions of the two possible twin orientations in relative amounts  $f^s$ ,  $1-f^s$ , the equivalent lattice-invariant shear is  $k_+^M = sf^s$  [or  $k_-^M = s(1-f^s)$ ] depending on which correspondence is regarded as giving the first ellipsoid. The two lattice strains can thus produce the same result as the combination of a lattice strain and a lattice-invariant strain; the interface is an invariant plane for vectors which cross a reasonable number of twin lamellae. Twinned arrangements of this kind are obviously effective only if  $f^s$  is between zero and one. There is nothing in the formal crystallography to prevent a solution in which  $k_+^M > s$ ,  $k_-^M < 0$  and both orientations could then be present in a single plate, but the shape requirements would then necessitate a slip shear, i.e. an additional lattice-invariant deformation. If the elements  $\mathbf{I}^M$  and  $\mathbf{m}^M$  have been selected solely because they correspond to observed twinning elements of the product structure, there may be some justification for rejecting solutions which cannot be obtained by twinning alone. However, this position was not taken by Bowles and Mackenzie, who supposed the shape requirements to be met by a lattice-invariant strain even though the elements of this strain are twinning elements of the product structure. In the early WLR papers the above results are implicitly assumed and the product is regarded as a stack of fine twins, the twinning direction and plane being selected intuitively.

The BB theory will be discussed next as the sequence of matrix operations is closely similar to that of the WLR theory. Bullough and Bilby lay emphasis on the formal concept of a surface dislocation which is a second-rank tensor specifying the net dislocation content of a semi-coherent interface (see Section 38). In a martensitic transformation, the condition that the boundary be mobile implies that it corresponds to a glissile surface dislocation, of which the simplest example is an array of parallel glide dislocations all having the same Burgers vector. The motion of this array produces a lattice-invariant shear in the direction of the Burgers vector, the glide plane of the dislocations being the invariant plane of the shear.

In this theory, it is convenient to use the principal axes of the lattice deformation to define an orthonormal basis in the parent lattice. It should be remembered that lattice vectors need not have integral components when referred to such a basis. The lattice deformation is now specified by

$$\mathbf{v} = \mathbf{R}\bar{\mathbf{P}}\mathbf{u} \quad (92.8)$$

where  $\bar{\mathbf{P}}$  is a diagonal representation of the pure lattice strain. The theory of surface dislocations states that if  $\mathbf{p}$  is any unit vector in the boundary between the phases, then the net Burgers vector of the dislocation lines cutting  $\mathbf{p}$  is  $(\bar{\mathbf{P}}^{-1}\mathbf{R}^{-1} - \mathbf{I})\mathbf{p}$  [see eqn. (38.2)]. For a simple glissile array, the Burgers vector must be parallel to the unit vector  $\mathbf{l}$  in the direction of the Burgers vector of the individual lines and may be written  $t\mathbf{l}$ . Hence

$$t\mathbf{l} = (\bar{\mathbf{P}}^{-1}\mathbf{R}^{-1} - \mathbf{I})\mathbf{p} \quad (92.9)$$

where  $t$  is a parameter which varies with the direction of  $\mathbf{p}$ .

Bullough and Bilby consider the slip shear to be applied to the parent lattice before the lattice deformation, so that

$$\mathbf{v} = \mathbf{R} \bar{\mathbf{P}} \mathbf{K} \mathbf{u} = \mathbf{E} \mathbf{u}. \quad (92.10)$$

Let the plane containing  $\mathbf{l}$  and any vector  $\mathbf{u}$  meet the habit plane in the unit vector  $\mathbf{p}$ . Then  $\mathbf{u}$  may be written  $\mathbf{u} = \alpha \mathbf{l} + \beta \mathbf{p}$ . As a result of the slip deformation, the component  $\alpha \mathbf{l}$  is unaffected but the end points of  $\beta \mathbf{p}$  are displaced by the net Burgers vector cut by  $\beta \mathbf{p}$ . Thus the new vector is  $\mathbf{K} \mathbf{u} = \alpha \mathbf{l} + \beta \mathbf{p} + \beta t \mathbf{l} u + \beta t \mathbf{l}$ . However,  $\mathbf{K}$  represents a simple shear on the plane  $\mathbf{m}$  in the direction  $\mathbf{l}$  and hence may be written

$$\mathbf{K} = \mathbf{I} + k \mathbf{l} \mathbf{m}' \quad (92.11)$$

where  $k$  specifies the amount of shear. The vector  $\mathbf{K} \mathbf{u}$  is thus  $\mathbf{u} + k \mathbf{l} \mathbf{m}' \mathbf{u} = \mathbf{u} + \beta k \mathbf{l} \mathbf{m}' \mathbf{p}$ . Comparing the two expressions for  $\mathbf{K} \mathbf{u}$  gives

$$t = k \mathbf{m}' \mathbf{p}$$

Now let  $\alpha = 0$  so that  $\mathbf{u} = \beta \mathbf{p}$  and  $\mathbf{K} \mathbf{u} = \beta \mathbf{p} + \beta t \mathbf{l} = \beta \bar{\mathbf{P}}^{-1} \mathbf{R}^{-1} \mathbf{p}$  from eqn. (92.9). Hence

$$\mathbf{E} \beta \mathbf{p} = \mathbf{R} \mathbf{P} \mathbf{K} \beta \mathbf{p} = \beta \mathbf{p} \quad (92.12)$$

Thus all vectors in the habit plane are unchanged by the total shape deformation when the dislocation content of the boundary is specified by eqn. (92.9).

The habit plane is defined by the unity principal axis of the reciprocal shape ellipsoid and by the invariant line of the whole deformation. One approach in the **BB** theory, as in the **WLR** theory, is to find the invariant line which we denote as  $\mathbf{d}$ . Clearly  $\mathbf{d}$  is in the direction of the dislocation lines if they form a single parallel array, and when  $\mathbf{p} = \mathbf{d}$ ,  $t = 0$  in eqn. (92.9). This gives  $\mathbf{R}' \mathbf{d} = \bar{\mathbf{P}} \mathbf{d}$  or, as  $\bar{\mathbf{P}}$  is diagonal,

$$\mathbf{d}' (\mathbf{I} - \bar{\mathbf{P}}^2) \mathbf{d} = 0 \quad (92.13)$$

When written out in full, this equation is

$$d_1^2 (1 - \eta_1^2) + d_2^2 (1 - \eta_2^2) + d_3^2 (1 - \eta_3^2) = 0 \quad (92.14)$$

and merely expresses the condition that  $\mathbf{d}$  lies in the cone of lines which do not change length during the lattice deformation. Equation (92.14) is obtained directly by substituting  $\bar{x}_i = d_i$  and  $\lambda_i = \eta_i$  in eqn. (9.15).

The components of  $\mathbf{d}$  must also satisfy the condition that  $\mathbf{d}$  is normal to  $\mathbf{m}$  and is a unit vector, i.e.

$$d_i m_i = 0 \quad (92.15)$$

$$d_1^2 + d_2^2 + d_3^2 = 1 \quad (92.16)$$

From eqns. (92.14), (92.15) and (92.16) a quadratic equation is obtained for each component of  $\mathbf{d}$ . Thus in general there are two possible invariant lines  $\mathbf{d}_+$  and  $\mathbf{d}_-$  for any choice of  $\mathbf{p}$  and  $\mathbf{m}$ ; these correspond to the intersections of  $\mathbf{m}$  with the cone represented by eqn. (92.14). The quadratic has real roots only if

$$m_1^2 (1 - \eta_2^2)(1 - \eta_3^2) + m_2^2 (1 - \eta_3^2)(1 - \eta_1^2) + m_3^2 (1 - \eta_1^2)(1 - \eta_2^2) \leq 0 \quad (92.17)$$

This inequality includes the restriction mentioned on p. 61 that all the  $(1 - \eta_i^2)$  should not have the same sign, otherwise the cone will not exist, and the restriction that the plane  $\mathbf{m}$  must intersect the cone. For a particular transformation in which  $\mathbf{P}$  is known, it may thus be regarded as restricting the choice of  $\mathbf{m}$ . Crocker (1959) points out that it is convenient to plot eqn. (92.17) on a stereogram (the “ $\mathbf{m}$  restriction curve”), no plane having a pole inside this curve being available as the invariant plane of the slip shear.

Multiplying both sides of eqn. (92.9) by  $\mathbf{p}$  and rearranging gives

$$\bar{\mathbf{P}}\mathbf{p} + t\bar{\mathbf{P}}\mathbf{I} = \mathbf{R}'\mathbf{p} \quad (92.18)$$

and multiplying on the left by  $\mathbf{d}'\bar{\mathbf{P}}^2$  and using  $\mathbf{d}'\mathbf{R} = \mathbf{d}'\mathbf{P}$  gives

$$t\mathbf{d}'\bar{\mathbf{P}}^2\mathbf{I} = \mathbf{d}'(\mathbf{I} - \bar{\mathbf{P}}^2)\mathbf{p} \quad (92.19)$$

If each side of eqn. (92.19) is now multiplied by its own transpose and  $t$  is eliminated by using eqn. (92.18), the apparently complex equation

$$[\mathbf{d}'\bar{\mathbf{P}}^2\mathbf{I}]^2[\mathbf{p}'(\mathbf{I} - \bar{\mathbf{P}}^2)\mathbf{p}] = 2[\mathbf{d}'\bar{\mathbf{P}}^2\mathbf{I}][\mathbf{p}'\bar{\mathbf{P}}^2\mathbf{I}][\mathbf{d}'(\mathbf{I} - \bar{\mathbf{P}}^2)\mathbf{p}] + [\mathbf{I}'\bar{\mathbf{P}}^2\mathbf{I}][\mathbf{d}'(\mathbf{I} - \bar{\mathbf{P}}^2)\mathbf{p}]^2, \quad (92.20)$$

in which each quantity in square brackets is a scalar product, is obtained.

This equation must be satisfied by all vectors  $\mathbf{p}$  in the habit plane and hence it implicitly defines the habit plane normal  $\mathbf{v}$ . To obtain an explicit solution, the components of  $\mathbf{p}$  must be eliminated by using the three equations

$$\begin{aligned} p_i v_i &= 0 \\ d_i v_i &= 0 \\ (1 - \eta_i^2) d_i^2 &= 0 \end{aligned} \quad (92.21)$$

The last of these equations is eqn. (92.14) and the other two follow because  $\mathbf{p}$  and  $\mathbf{d}$  are both in the habit plane. The resultant algebra is rather lengthy, but use of eqn. (92.19) with eqn. (92.21) eventually leads to the following quadratic equation

$$L_i v_i^2 + M v_i v_j + L_j v_j^2 = 0 \quad (92.22)$$

where

$$\begin{aligned} L_k &= N_{(k)}[N_{(k)}P + 2T_{(k)}Q - Q^2/d_{(k)}] \\ M &= -2[N_i N_j P + (N_j T_i + N_i T_j)Q] \\ N_k &= d_{(k)}(1 - \eta_{(k)}^2), \quad T_{(k)} = 1_{(k)}\eta_{(k)}^2 \\ P &= \sum_{k=1}^3 l_k^2 n_k^2, \quad Q = \sum_{k=1}^3 l_k d_k \eta_k^2 \end{aligned} \quad (92.23)$$

and the parentheses around the subscripts indicate that the summation convention is not used.

Equation (92.23) gives directly the ratio of any two components of  $\mathbf{v}$ ; the third component may be obtained by solving the equation twice or, more simply, from eqn. (92.21). The components are determined completely by the normalizing condition

needed to make  $\mathbf{v}$  a unit vector. All the quantities in eqn. (92.22) are known from the initial assumptions once eqns. (92.14), (92.15) and (92.16) have been solved for the components of  $\mathbf{d}$ . Thus this procedure gives the habit plane by solving consecutively two quadratic equations.

In the simpler applications of the theory, it is possible to obtain parametric expressions for the habit plane in terms of  $\eta_i^2$  without following the above treatment of the BB theory. However, this solution has been developed in its most general form as this gives an especially convenient method of varying the initial assumptions about the lattice-invariant deformation. Crocker and Bilby (1961a) programmed the solution for a computer and obtained thousands of habit plane predictions for a single transformation (see Section 92).

For each of the two values of  $\mathbf{d}$ , two values of  $\mathbf{v}$  will be obtained, corresponding to the sign of the root in the quadratic. Thus there are in general four different solutions for the habit plane, as concluded above. The choice of sign in the quadratic for  $\mathbf{d}$  corresponds to the choice of sign in eqn. (91.8) as is evident geometrically from Fig. 22.2. The choice of sign in the quadratic (92.22) must then correspond to the two possible values of the lattice-invariant shear.

The condition that any one of the three quadratics represented by eqn. (92.23) shall have a real solution now leads to an  $\mathbf{l}$  restriction similar to the  $\mathbf{m}$  restriction above in which each  $\mathbf{m}_i$  in eqn. (92.17) is replaced by  $\mathbf{l}_i$ . For given  $\mathbf{P}$ , this restricts the choice of shear direction and it may also be plotted stereographically. The pure lattice deformation  $\mathbf{P}$  gives a cone of lines of unchanging length and also a cone of normals to planes of unchanging spacing. All planes containing  $\mathbf{l}$  will have spacings unchanged by the lattice-invariant deformation. Hence after both lattice and lattice-invariant deformation, there will be one set of planes of unchanging spacing; the normal to these planes will be a line in the lattice deformation cone of plane normals which is itself normal to  $\mathbf{l}$ . The existence condition for  $\mathbf{l}$  is that the plane normal to  $\mathbf{l}$  shall intersect the cone of plane normals of unchanging length.

The intersection of the unit sphere and the reciprocal strain ellipsoid (for the lattice deformation) gives the cone of vectors of unchanging length in its initial position and the cone of plane normals of unchanging length after the pure deformation  $\mathbf{P}$  (i.e. relative to the product lattice). The intersection of the unit sphere and the first ellipsoid gives the directions of unchanging length after the pure lattice deformation, or to normals to planes of unchanging spacing in their initial positions. Thus the  $\mathbf{m}$  and  $\mathbf{l}$  restrictions together express the conditions that the invariant plane of the shear and the plane normal to the shear direction must intersect the first and second of these cones respectively when  $\mathbf{m}$  and  $\mathbf{l}$  are referred to the parent lattice and the second and first of these cones when referred to the product lattice.

From eqn. (92.17), the magnitude of the lattice-invariant shear is given by

$$k = \frac{\mathbf{d}'(\mathbf{I} - \bar{\mathbf{P}}^2)\mathbf{p}}{Q\mathbf{m}'\mathbf{p}} \quad (92.24)$$

and using  $\mathbf{p}'\mathbf{v} = \mathbf{d}'\mathbf{v} = \mathbf{m}'\mathbf{d} = 0$ , this becomes

$$k = \frac{v_i d_{(i)}(1 - \eta_{(i)}^2) - v_j d_{(j)}(1 - \eta_{(j)}^2)}{Q(m_j v_i - m_i v_j)} \quad (92.25)$$

so that there is a single value of  $k$  for each habit plane. The equation apparently implies that there are four values of  $k$  from one set of input data but substitution for  $\mathbf{d}$  and  $\mathbf{v}$  leads to a rather complex quadratic equation, which corresponds to the equation for  $\mathbf{l}^M$  derived from eqn. (92.4). There are thus only two values of  $k$  as deduced earlier.

Expression (92.26) gives the lattice-invariant shear in the parent lattice. In accordance with the BB procedure, whilst eqn. (92.4) gives the shear,  $k^M$  gives the corresponding shear in the product lattice. The two are related by

$$k^M = |\mathbf{PI}| |\mathbf{m}' \bar{\mathbf{P}}^{-1}| k \quad (92.26)$$

From  $k$  and eqn. (92.11), it is possible to write down explicitly the matrix representation of the lattice-invariant deformation  $\mathbf{K}$ . To complete the theory and specify the orientation relations and the shape deformation, the matrix  $\mathbf{R}$  specifying the rotational part of  $\mathbf{S}$  must be found. The most convenient way of doing this in the general case was first given by Lieberman *et al.* (1955). Equation (92.13) shows that for any vector in the habit plane

$$\bar{\mathbf{P}}\mathbf{K}\mathbf{p} = \mathbf{R}'\mathbf{p}$$

i.e., as is geometrically obvious, the rotation produced by  $\mathbf{R}$  is equal and opposite to that produced by  $\mathbf{PK}$  for habit plane vectors. In general  $\mathbf{R}$  represents a right-handed rotation about some general vector  $\mathbf{r}$  through an angle  $\theta$ . There is then a standard result that if  $\mathbf{s}$  and  $\mathbf{t}$  are any two non-parallel vectors, not parallel to  $\mathbf{r}$ ,

$$\left( \tan \frac{1}{2} \theta \right) \mathbf{r} = \frac{(\mathbf{s} - \mathbf{R}'\mathbf{s}) \wedge (\mathbf{t} - \mathbf{R}'\mathbf{t})}{(\mathbf{s} - \mathbf{R}'\mathbf{s}) \cdot (\mathbf{t} - \mathbf{R}'\mathbf{t})} \quad (92.27)$$

Hence  $\mathbf{r}$  and  $\theta$  may be found by choosing  $\mathbf{s}$  and  $\mathbf{t}$  to be two vectors in the habit plane. Convenient choices are  $\mathbf{s} = \mathbf{d}$  and  $\mathbf{t} = \mathbf{v} \wedge \mathbf{d}$ . In terms of  $\theta$  and  $\mathbf{r}$ , the elements of the matrix  $\mathbf{R}$  are

$$R_{ij} = \delta_{ij} \cos \theta + r_i r_j (1 - \cos \theta) - \varepsilon_{ijk} r_k \sin \theta \quad (92.28)$$

Although in principle these matrix elements could be written in parametric form, this is not feasible in practice.

The matrix elements of the shape deformation  $\mathbf{E}$  in the orthonormal system may now be written down from the elements of  $\mathbf{K}$ ,  $\mathbf{P}$  and  $\mathbf{R}$ . The shape deformation changes any macroscopic vector  $\mathbf{u}$  into  $\mathbf{E}\mathbf{u}$  and the vector  $\mathbf{E}\mathbf{u} - \mathbf{u}$  thus gives the direction of displacement, which lies in the macroscopic habit if there is no volume change in the transformation. The shape deformation is illustrated in Fig. 22.3; it is completely specified by the habit plane and by the displacement  $C_{31}\mathbf{e} = \mathbf{E}\mathbf{v} - \mathbf{v}$  of a point at unit distance from a reference habit plane. The unit vector  $\mathbf{e}$  defines the direction of the displacement and  $C_{31}$  measures the magnitude of the shape change. A more convenient measure of the

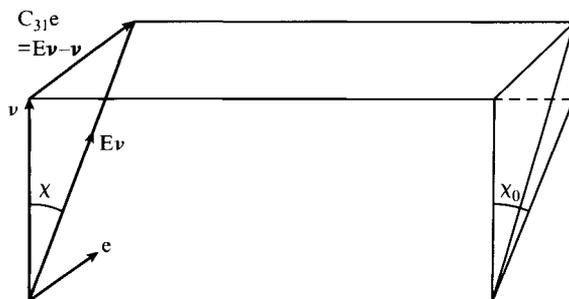


FIG. 22.3. Parameters used in defining the shape deformation.  $v$  is the normal to the habit plane,  $e$  is a unit vector in the direction of the displacement and  $C_{31}$  defines the magnitude of the shape change.

magnitude is the angle  $\chi$  between the initial and final positions of a vector originally normal to the habit plane (see Fig. 22.3). This angle is given by

$$\chi = \cos^{-1} \left( \frac{v \cdot E v}{|E v|} \right) \tag{92.29}$$

An experimental quantity more often measured is the angle  $\chi^0$  between  $v$  and the projection of  $C_{31} v$  on to the habit plane; this defines the apparent macroscopic shear.

There is no generally agreed method of presentation of the orientation relations. Experimental work is usually reported as the angles between nearly parallel rational directions and planes of the two structures, but this is not very satisfactory and the statements sometimes contain inconsistencies. A better specification is to report the direction  $r$  and angle  $\theta$  which represent the relation between orthonormal bases defined in the two structures, or the deviation from some classic rational orientation which can be defined in terms of the exact parallelism of various directions and/or planes. For example, in the important f.c.c.–b.c.t. transformations, a useful reference orientation is the Kurdjumov–Sachs orientation in which the closest-packed planes of the two structures are parallel, as are the closest-packed directions within these planes.

An alternative way of representing the relation between bases or the deviation from a reference orientation is to use the Euler angles (Fig. 22.4). Here, if the two orthonormal bases are  $I = i_i, J = j_j$ ,  $\theta$  is the angle between  $i_3$  and  $j_3$ ,  $\phi$  is the angle between the plane containing  $i_3$  and  $j_3$  and the plane containing  $i_1$  and  $i_3$ , and  $\psi$  is the angle between the plane containing  $i_3$  and  $j_3$  and the plane containing  $j_1$  and  $j_3$ .

The BM theory follows the traditions of the earlier Greninger–Troiano analysis in working mainly with lattice deformation rather than shape deformations. The alternative factorizations of  $S$  may be written

$$S = M E = E S^{-1} M S \tag{92.30}$$

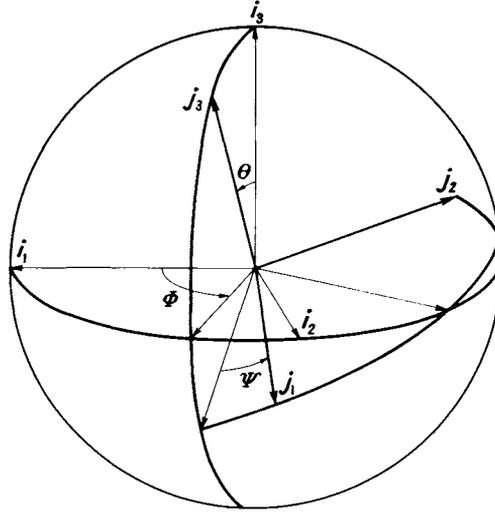


FIG. 22.4. The Eulerian angles used to specify a rigid body rotation between the axes  $i_1 i_2 i_3$  and  $j_1 j_2 j_3$ .

where  $\mathbf{M}$  and  $\mathbf{S}^{-1}\mathbf{M}\mathbf{S}$  represent the invisible part of the lattice deformation applied respectively before and after the visible part. We can write  $\mathbf{S}^{-1}\mathbf{M}\mathbf{S} = \mathbf{K}^{-1}$  and the resulting equation

$$\mathbf{S} = \mathbf{E}\mathbf{K}^{-1}$$

is equivalent to the BB formulation [eqn. (92.10)]. Bowles and Mackenzie introduced a dilatation parameter in their analysis and this has the effect of introducing an additional term  $\delta$  into the shape deformation. The total lattice deformation, now called  $\mathbf{S}_T$ , may be imagined to occur in two stages, the first of which is the formation of the smaller (or larger) sphere of radius  $1/\delta$ . The deformation carrying this sphere into the first ellipsoid is the invariant line strain  $\mathbf{S}$  which is to be factorized. This means that the habit plane will contain one of the directions in the cone  $POQ$  (see Fig. 22.5) rather than one of the lines of unchanging length  $ROS$ .

Bowles and Mackenzie introduced an orthonormal basis  $\mathbf{l}$  defined by  $\mathbf{i}_1 = \mathbf{l} \wedge \mathbf{m}$ ,  $\mathbf{i}_2 = \mathbf{l}$ ,  $\mathbf{i}_3 = \mathbf{m}$ . In this basis, the equation of the reciprocal strain ellipsoid (for the deformation  $\mathbf{S}$ ) is

$$(\mathbf{x}; \mathbf{l})(|\mathbf{S}\mathbf{l}|)'(|\mathbf{S}\mathbf{l}|)[\mathbf{l}; \mathbf{x}] = (\mathbf{x}; \mathbf{l})[\mathbf{l}; \mathbf{x}]$$

or in scalar form

$$H_{11}x_1^2 + H_{22}x_2^2 + H_{33}x_3^2 + 2H_{12}x_1x_2 + 2H_{23}x_2x_3 + 2H_{31}x_3x_1 = 1 \tag{92.31}$$

where  $H_{ij}$  are the elements of the symmetric matrix

$$\mathbf{H} = (|\mathbf{S}\mathbf{l}|)(|\mathbf{S}\mathbf{l}|)' = (|\mathbf{l}|)(|\mathbf{l}|)' \tag{92.32}$$

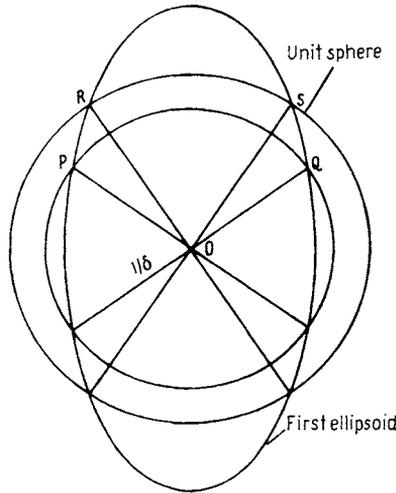


FIG. 22.5. The geometry of the Bowles-Mackenzie theory. All lines in the cone  $ROS$  (shown in section) are unchanged in length; all lines in the cone  $POQ$  change in length by  $1/\delta$ . The habit plane contains one of the lines in  $POQ$ . Note that  $POQ$  could exist when  $ROS$  is imaginary (all principal strains have the same sign), but in all known transformations  $ROS$  is real.

The corresponding coefficients of the equation of the first ellipsoid in the basis  $\mathbf{l}$  are written  $H_{ij}^*$  and form the elements of  $\mathbf{H}^{-1}$ . All the elements of  $\mathbf{H}$  and  $\mathbf{H}^{-1}$  are known when  $\mathbf{P}$ ,  $\mathbf{m}$  and  $\mathbf{l}$  are fixed. We now use the additional notation  $H = |\mathbf{H}|^{1/2}$  and  $H^* = |\mathbf{H}^{-1}|^{1/2}$ . (This special notation is used only for the matrix  $\mathbf{H}$  and its reciprocal. It is introduced mainly to facilitate comparison with the BM papers. They call the matrices  $\mathbf{G}$  and  $\mathbf{G}^{-1}$  and the elements  $g_{ij}$ ,  $g_{ij}^*$ , but  $\mathbf{H}$  is preferred here to avoid possible confusion with the metric tensor.) The ellipse in which the plane  $\mathbf{m}$  meets the reciprocal strain ellipsoid has the equation

$$H_{11}x_1^2 + H_{22}x_2^2 + 2H_{12}x_1x_2 = 1 \tag{92.33}$$

and is shown in Fig. 22.6. The lines of unchanging length  $\mathbf{d}_+$  and  $\mathbf{d}_-$  are given by the intersection of this ellipse with the unit circle  $x_1^2 + x_2^2 = 1$  and have equations

$$x_2 = x_1 \tan \alpha \tag{92.34}$$

where  $(H_{22} - 1)\tan^2 \alpha + 2H_{12} \tan \alpha + H_{11} - 1 = 0$   
i.e.

$$(H_{22} - 1)\tan \alpha = -H_{12} \pm \{H_{12}^2 - (H_{11} - 1)(H_{22} - 1)\}^{1/2} \tag{92.35}$$

The two values  $\alpha_1$  and  $\alpha_2$  given by this equation are used to characterize the two possible invariant lines; the condition that it has real roots is given by eqn. (92.17). The two values of  $\alpha$  are different in general but become equal and opposite when  $H_{12} = 0$ , i.e. when either  $i_1$  or  $i_2$  is parallel to a principal axis of the reciprocal strain ellipsoid. If this axis is also a twofold symmetry axis, the two lines  $\mathbf{d}_+$  and  $\mathbf{d}_-$  are crystallographically equivalent. This is

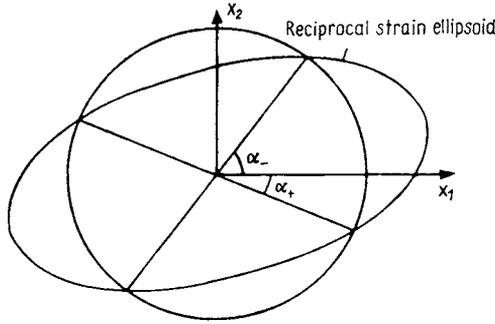


FIG. 22.6. Section of reciprocal strain ellipsoid by the plane  $\mathbf{m}$ . The lines of unchanging length make angles  $\alpha_+$ ,  $\alpha_-$  with the  $x_1$  axis.

the conclusion reached on p. 998, the two values of  $\alpha$  corresponding to the choice of sign in eqn. (92.6).

The initial positions of the two possible invariant normals are similarly given by

$$x_2 = x_1 \tan \beta \tag{92.36}$$

in the basis  $\mathbf{l}$ , where

$$(H_{33}^* - 1) \tan \beta = -H_{13}^* \pm \{H_{13}^{*2} - (H_{11}^* - 1)(H_{33}^* - 1)\}^{1/2} \tag{92.37}$$

The four possible invariant line strains are now each defined by one value of  $\alpha$  and one of  $\beta$ , but are rather difficult to calculate in this way. An alternative procedure is to fix one invariant line by selecting one of the  $\alpha$  values, and then to consider the invariant line strains which differ only by rotations about this invariant line. In particular one such strain, called here  $\mathbf{S}'$ , will leave the plane normal  $\mathbf{m}$  unrotated. Having calculated  $\mathbf{S}'$ , the further rotation needed to give  $\mathbf{S}$ , which leaves  $\nu$  unrotated, can be calculated.

A new orthonormal system  $\mathbf{K}$  is now introduced with  $\mathbf{k}_1 = \mathbf{d}$ ,  $\mathbf{k}_2 = \mathbf{m} \wedge \mathbf{d}$  and  $\mathbf{k}_3 = \mathbf{m}$ . The change of axes is merely a rotation of  $\alpha$  about  $\mathbf{m}$ . The deformation  $\mathbf{S}'$  has to leave the direction  $\mathbf{k}_1$  invariant and the normal to the plane containing  $\mathbf{k}_1$  and  $\mathbf{k}_2$  unrotated. The matrix representation  $\mathbf{S}'$  thus has components  $S'_{11} = 1$ ,  $S'_{21} = S'_{31} = S'_{32} = 0$ . The remaining components are obtained from the condition that the principal axes are those of  $\mathbf{P}$  and hence are known relative to the  $\kappa$  basis. In fact

$$(\mathbf{KS}'\mathbf{K})'(\mathbf{KS}'\mathbf{K}) = (\mathbf{KPK})'(\mathbf{KPK}) = (\mathbf{KSK})'(\mathbf{KSK}) \tag{92.38}$$

or

$$(\mathbf{KS}'\mathbf{K})'(\mathbf{KS}'\mathbf{K}) = (\mathbf{KLI})\mathbf{H}(\mathbf{ILK})$$

where

$$(\mathbf{ILK}) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

By equating like coefficients on the two sides, the unknown  $S_{ij}^o$  are determined. These equations are obtained in the most symmetrical form by considering also the inverse relations

$$(\mathbf{KS}^o\mathbf{K})^{-1}(\mathbf{KS}^o\mathbf{K})'^{-1} = (\mathbf{KL})\mathbf{H}^{-1}(\mathbf{ILK}) \tag{92.39}$$

Equations (92.38) and (92.39) together give  $S_{ij}^o$  in terms of  $H_{ij}$  and  $H_{ij}^*$ .

$\mathbf{S}$  differs from  $\mathbf{S}^o$  by a rotation  $\mathbf{R}^o$  about  $\mathbf{k}_1 = \mathbf{d}$ . If the amount of this rotation is  $\omega$ , then

$$(\mathbf{KR}^o\mathbf{K}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & -\sin \omega \\ 0 & \sin \omega & \cos \omega \end{pmatrix} \tag{92.40}$$

The angle  $\omega$  is determined by the condition that the plane normal to  $\mathbf{l}$  contains the invariant normal. As  $[\kappa; \mathbf{l}] = [\sin \alpha, \cos \alpha, 0]$

$$(\mathbf{KSK})[\mathbf{K}; \mathbf{l}] = (\mathbf{KR}^o\mathbf{K})(\mathbf{KS}^o\mathbf{K})[\mathbf{K}; \mathbf{l}]$$

in terms of the unknown  $\omega$ . The plane defined by  $[\mathbf{K}; \mathbf{l}]$  and  $(\mathbf{KSK})[\mathbf{K}; \mathbf{l}]$  has a normal

$$[S_{22}^o \sin \omega \cos \alpha, S_{22}^o \sin \omega \sin \alpha, (S_{22}^o \cos \omega - 1) \sin \alpha - S_{12}^o \cos \alpha]$$

in the  $\mathbf{K}$  basis. The condition that this be the invariant normal of the deformation  $\mathbf{S}$  leads to an equation for  $\omega$  in terms of  $S_{ij}^o$  and hence of the elements of  $\mathbf{H}$  and  $\mathbf{H}^{-1}$ . This equation is

$$\begin{aligned} & HH_{13}^* \sin \omega + [(HH_{33}^* + H_{22}) \sin \alpha + H_{12} \cos \alpha] \cos \omega \\ & = H_{33}^{*1/2} [(H + H_{22}) \sin \alpha + H_{12} \cos \alpha] \end{aligned} \tag{92.41}$$

and it is satisfied by two distinct values  $\omega_+$  and  $\omega_-$ . These two values replace the two  $\beta$  values above, and possible strains  $\mathbf{S}$  are obtained by choosing one value of  $\alpha$  and one of  $\omega$ .

Provided the  $\mathbf{l}$  condition is satisfied,  $\mathbf{l}$  can be any direction in  $\mathbf{m}$ . The special assumption made in the original formulation of the **BM** theory was, however, that  $\mathbf{l}$  and  $\mathbf{m}$  are twinning elements of the product lattice and that the correspondences of the two possible twin lattices are crystallographically equivalent variants. It has already been proved (p. 999) that for type I twinning  $\mathbf{m}$  must be a mirror plane of the parent structure, and it will now be shown that  $\mathbf{l}$  is fixed once  $\mathbf{m}$  is chosen. In Fig. 22.7, let  $O$  and  $O'$  be reflections of each other in the mirror plane  $ABC$  of the parent structure. Use any orthonormal basis in which the vector  $OO'$  is represented by the column matrix  $2C_{20}\mathbf{m}$ , where  $\mathbf{m}$  is a unit vector normal to  $ABC$  and  $C_{20}$  is the distance of  $O$  from  $ABC$ . After the pure lattice deformation,  $O'$  will be displaced to  $O'''$ , where  $O'''$  is represented by  $2C_{20}\mathbf{Pm}$ . The plane  $ABC$  becomes  $A'B'C'$  with normal  $\mathbf{m}^M = \mathbf{P}^{-1}\mathbf{m}$ . As  $\mathbf{P}$  is symmetric and the distance of  $O$  from the plane is  $C_{20}/|\mathbf{m}^M|$ . The image of  $O$  in  $A'B'C'$  is now  $O''$ , where  $OO''$  is represented by  $2C_{20}\mathbf{P}^{-1}\mathbf{m}/|\mathbf{m}^M|^2 = 2C_{20}\mathbf{P}^{-1}\mathbf{m}/\mathbf{m}'\mathbf{P}^{-2}\mathbf{m}$ . If  $A'B'C'$  is a twinning plane of the product structure, the twin orientation will have a lattice point at  $O''$  and hence the twinning shear displaces  $O'''$  to  $O''$ . The twinning ( $\eta_2$ ) direction in the product lattice is thus parallel to the vector  $O''' - O'' = OO' - O'''$ . When referred to the original orthonormal basis

$$O'''O'' = 2C_{20}[(\mathbf{P}^{-1}\mathbf{m}/\mathbf{m}'\mathbf{P}^{-2}\mathbf{m}) - \mathbf{Pm}]$$

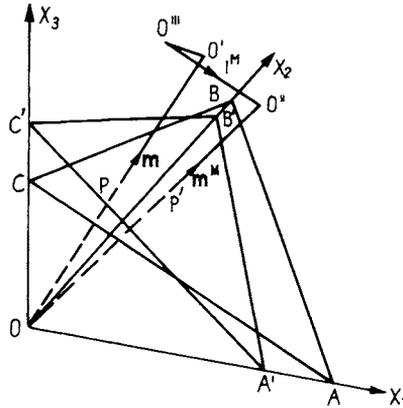


FIG. 22.7. Determination of the twinning direction  $\mathbf{l}$  for given  $\mathbf{m}$ .  $OP$  (parallel to  $\mathbf{m}$ ) is the perpendicular from the origin to the plane  $ABC$ . After the pure deformation the plane becomes  $A'B'C'$  and the perpendicular is  $OP'$ , parallel to  $\mathbf{m}^M$ . The images of  $O$  in  $ABC$  and  $A'B'C'$  are  $O'$  and  $O''$ ; the pure strain displaces the point  $O'$  to  $O'''$ .  $\mathbf{l}^M$  is parallel to  $O'''O''$ .

The magnitude of the twinning shear is

$$s = |O'''O''| |\mathbf{P}^{-1}\mathbf{m}| / C_{20} \quad (92.42)$$

Let the displacement vector  $O'''O''$  be derived from the corresponding vector of the parent lattice  $O'O''''$  so that  $O''''$  is the point which moves to  $O''$  after the deformation  $\mathbf{P}$ . Then

$$O'O'''' = OO'''' - OO' = 2C_{20}[(\mathbf{P}^{-2}\mathbf{m}/\mathbf{m}'\mathbf{P}^{-2}\mathbf{m}) - \mathbf{m}]$$

The twinning shear in the parent lattice is represented by the vector  $s^p\mathbf{l}$  where

$$s^p\mathbf{l} = 2[(\mathbf{P}^{-2}\mathbf{m}/\mathbf{m}'\mathbf{P}^{-2}\mathbf{m}) - \mathbf{m}] \quad (92.43)$$

This is the result obtained by Bowles and Mackenzie and it shows that the direction  $\mathbf{l}$  of the twinning shear is fixed by the choice of  $\mathbf{m}$  if it is assumed that the two twin orientations have equivalent lattice correspondence with the parent phase.

The result just derived applies to all compound and type I twins which must be derived from a mirror plane of the parent. In some martensites, the twins are type II. It can then be shown in a similar manner that the direction of the twinning shear must be derived from a twofold axis of the parent structure and that, once  $\mathbf{l}$  is chosen, the plane  $\mathbf{m}$  is automatically fixed by the assumption that the twins have equivalent correspondence.

We now use a new basis  $\mathbf{M}$  in the product which has base vectors generated from those of the parent by the deformation  $\mathbf{P}$  (i.e. the correspondence matrix is  $\mathbf{l}$ ). In this basis, the displacement vector  $O'''O''$  of the product is

$$[O'''O'']_{\mathbf{M}} = 2C_{20}[(\mathbf{P}^{-2}\mathbf{m}/\mathbf{m}'\mathbf{P}^{-2}\mathbf{m}) - \mathbf{m}] = C_{20}s^p\mathbf{l} \quad (92.44)$$

and has of course the same representation as  $O'O''''$  in the parent phase. In the basis  $\mathbf{m}$ , the vector  $\mathbf{l}$  is not a unit vector but a vector of magnitude  $|\mathbf{Pl}|$ . Thus, from eqn. (92.42), the twinning shear in the product has magnitude

$$s = |\mathbf{Pl}||\mathbf{P}^{-1}\mathbf{m}|s^P \tag{92.45}$$

This is the relation between shear magnitudes in parent and product already used in eqn. (92.26).

Use of the BM twinning hypothesis leads to an automatic degeneracy in the results. In the basis  $\mathbf{l}$  (p. 1006), the vector  $\mathbf{P}^{-2}\mathbf{m}$  has components  $H_{13}^*$ ,  $H_{23}^*$  and  $H_{33}^*$  and  $\mathbf{m}'\mathbf{P}^{-2}\mathbf{m} = H_{33}^*$ . Hence from eqn. (92.42) the vector  $s^P\mathbf{l}$  is represented by the column matrix  $2[H_{13}^*/H_{33}^*$ ,  $H_{23}^*/H_{33}^*$ ,  $0]$ . But  $\mathbf{l} = \mathbf{i}_2$ , so that  $H_{13}^* = 0$  and  $s^P = 2H_{23}^*/H_{33}^*$ . From eqn. (92.37) it is evident that the two values of  $\beta$  are equal and opposite and the same result applies to the values of  $\omega$ . After some algebraic manipulation, eqn. (92.41) for  $\omega$  may be written

$$\sin \omega = \pm \frac{HH_{33}^{*1/2}(H_{33}^* - 1)(1 - H_{11}^*)^{1/2}}{(HH_{33}^* + H_{22}) \sin \alpha + H_{12} \cos \alpha} \tag{92.46}$$

The two values of  $\omega$  with any one value of  $\alpha$  give two crystallographically equivalent solutions, being related by a 180° rotation about  $\mathbf{m}$  as previously described.

All the components of the representation  $(\mathbf{KSK}) = (\mathbf{KR}^*\mathbf{K})(\mathbf{KS}^*\mathbf{K})$  may now be written down in terms of the known quantities  $H_{ij}$ ,  $H_{ij}^*$ ,  $\alpha$  and  $\omega$ . It remains to factorize  $\mathbf{S}$  as in eqn. (92.30).

During the macroscopic shape deformation  $\mathbf{E}$ , the plane normal represented by  $\mathbf{m}'$  becomes  $\mathbf{m}'\mathbf{E}^{-1}$  and the unit vector  $\mathbf{l}$  becomes  $\mathbf{El}$ . As these elements are invariant during the lattice-invariant deformation, it follows that  $\mathbf{Sl} = \mathbf{El}$  and  $\mathbf{m}'\mathbf{S}^{-1} = \mathbf{m}'\mathbf{E}^{-1}$ . Now because  $\mathbf{E}$  can be written as  $\mathbf{l} + C_{31}\mathbf{e}\mathbf{v}'$  it follows that  $C_{31}\mathbf{e}\mathbf{v}'\mathbf{l} = \mathbf{Sl} - \mathbf{l}$  and the direction of displacement  $\mathbf{e}$  is parallel to  $\mathbf{Sl} - \mathbf{l}$ . Similarly, the habit plane normal  $\mathbf{v}$  is parallel to  $\mathbf{m}'\mathbf{S}^{-1} - \mathbf{m}'$ . Using the basis  $\mathbf{K}$  and substituting for  $\mathbf{S} = (\mathbf{KSK})$

$$\begin{aligned} C_{32}(\mathbf{v}; \mathbf{K}) &= (0, \sin \omega, H_{33}^{* -1/2} - \cos \omega) \\ C_{33}(\mathbf{K}; \mathbf{e}) &= ((H_{22} - 1) \tan \alpha + H_{12}, HH_{33}^{*1/2} \cos \omega - 1, HH_{33}^{*1/2} \sin \omega) \end{aligned} \tag{92.47}$$

These equations give the unit vectors  $\mathbf{v}$  and  $\mathbf{e}$  corresponding to the habit plane normal and direction of displacement in the shape deformation; they may readily be referred to the basis  $\mathbf{l}$  by using the matrix (92.38). The magnitude of the shape deformation  $C_{31}$  is obtained from the condition  $\mathbf{Sl} = \mathbf{El}$  and is

$$C_{31} = C_{32}C_{33} \operatorname{cosec} \omega$$

An explicit expression for the lattice-invariant shear is given by

$$\mathbf{S}^{-1}\mathbf{MS} = \mathbf{l} - \mathbf{klm}' = \mathbf{l} - f^s s^P \mathbf{l}\mathbf{m}' \tag{92.48}$$

where  $f^s$  as before specifies the fraction of the twinning shear which is involved. This is a shear on the plane  $\mathbf{m}$  in the direction  $-\mathbf{l}$  and it is annulled by the twinning shear on the plane  $\mathbf{m}$  in the direction  $+\mathbf{l}$ . For any one value of  $\alpha$ , there are two values  $f_+^s$  and  $f_-^s$  corresponding to the two values of  $\omega$  such that

$$f_+^s + f_-^s = 1 \tag{92.49}$$

The corresponding values for the other value of  $\alpha$  are  $f_-^s$  and  $f_+^s$  respectively. The quantity  $1 - 2f_-^s$  changes only in sign for all four solutions and is given by

$$1 - 2f_-^s = \frac{\sin \omega [(H_{22} - 1) \tan \alpha + H_{12}] (H_{33}^* - 1)^{1/2}}{HH_{23}^* (1 - H_{11}^*)^{1/2}} \quad (92.50)$$

which is the equivalent of the earlier equations giving the amount of lattice-invariant shear.

The BM theory has now been developed to the point where a computing procedure can be set up for any particular transformation. The elements of  $H$  and  $H^{-1}$  are first computed [eqn. (92.32)], then  $\alpha$  is found with eqn. (92.35) and  $\omega$  with eqn. (92.46), giving  $\mathbf{v}$  and  $\mathbf{e}$  and also (KSK), (K RK) and (KEK).

When used in conjunction with the assumption of a dilatation, the results are obtained as a function of this unknown parameter, so that each predicted habit plane becomes a line on the stereographic projection showing how the prediction varies with the dilatation. For any one set of input data there are two non-equivalent predictions if the twinning hypothesis is used in the more general case. The number of possible assumptions for  $\mathbf{e}$  and  $\mathbf{m}$  is limited. In particular, if the structure is assumed to be twinned, and the twins are of type I,  $\mathbf{m}$  must be normal to a plane of type  $\{100\}$  or  $\{110\}$  even in cubic structures.

In conclusion, it should be emphasized again that, despite the different formulations, the theories are all equivalent and the special assumptions made in one or other of them may be readily discarded or incorporated into another description. Thus the description of the BM theory did not include the twinning hypothesis until p. 1009 and the computation process is equally valid when  $\mathbf{m}$  and  $\mathbf{l}$  are not restricted in the above manner. Equally, the calculations used in the WLR and BB theories may readily be adapted to the introduction of a dilatation parameter. The notation of this section has been chosen to facilitate a direct comparison with the original papers wherever possible, but the multiplicity of symbols makes this difficult. For convenience, the tensor and matrix notation used here is compared in Table XXI with that used in the original papers, and with that used by Wayman *et al.* (1961).

### 93. THE MARTENSITIC TRANSFORMATION IN STEELS

In carbon steels, martensite is the name given to the b.c.t. structure formed when the f.c.c. solid solution (austenite) is rapidly cooled. This phase may be regarded as a supersaturated solid solution of carbon in the low temperature, b.c.c. form of iron (ferrite), the carbon atoms occupying ordered octahedral positions as shown in Fig. 22.8 (Petch, 1943). Martensitic transformations are found in other iron alloys, the product being tetragonal when interstitial solutes are present and b.c.c. when only substitutional solutes are included. The transformation f.c.c.–b.c.c. in pure iron may also be martensitic in certain circumstances. The crystallography of all these changes will be described together and the bases of the conventional cubic and tetragonal unit cells will

TABLE XXI. COMPARISON OF NOTATION IN PAPERS ON MARTENSITE

Meaning	This book	W.L.R. (1953)	B.M (1954)	B.B. (1956)	Christian (1955-6)	Wayman (1961)
(Total) Lattice deformation	$S = ME$	$M_1M_2$	$S(S_T)$	<b>D</b>		<b>I</b>
Pure lattice deformation	<b>P</b>	$T_1T_2$	<b>M</b>			<b>B</b>
Pure deformation in principal axes system	<b>P</b>			<b>P</b>		<b>B</b>
Principal axes (co-ordinates) of lattice deformation	$i_i$	$i'j'k' \dots i''j''k''$	$p_i$	$x_i$	$(x_i)$	
Principal deformations of lattice	$\eta_i$	$\eta_i$	$\mu_i$	$e_i$	$\eta_i$	
Shear or twinning plane in initial position (and after lattice pure strain)	$m (m^M)$		<b>h</b>	<b>m</b>	<b>h (h')</b>	$\eta_i$
Shear or twinning direction in initial position (and after lattice pure strain)	<b>l (l<sup>M</sup>)</b>		<b>u</b>	<b>l</b>	<b>u (u')</b>	
Orthonormal axes (co-ordinates)	$i, k_i$		$i, l_i$		$(y_i)(z_i)$	
Lattice invariant deformation in parent (martensite)	<b>K (K<sup>M</sup>)</b>	<b>(G)</b>		<b>S</b>		<b>P</b>
Invisible lattice deformation	<b>M</b>		<b>P<sub>2</sub></b>			
Amount of shear in parent (martensite)	$k (k^M)$	$(g)$		$g (G)$	$(g)$	
Relative volumes of twins or fraction of twinning shear	$f, 1-f$	$x, 1-x, f, 1-f$	$k_{+0}, k_{-0}$			
Pure rotations	<b>R</b>	$\varphi_1\varphi_2$	<b>R</b>	<b>R</b>		<b>R</b>
Shape deformation	$E = RF = RPK$	$E = \varphi_1E = \varphi_1GT_1$	$P_1/\delta$	$F = RPS$		$P_1 = RnP$
Principal shape deformations	$\lambda_i$	$\lambda_i$				$\lambda_i$
Principal axes (co-ordinates) of shape deformation	$j_i(y_i)$	$(x_d, y_d, z_d)$			$(x_d, y_d, z_d)$	
Invariant line of lattice strain, after removing any dilatation	$d, d$	<b>d</b>	$l_1$	$n, n$		
Normal to planes of invariant spacing	$n, n$		<b>n</b>			
Habit plane normal	<b>v</b>	<b>n</b>	<b>p</b>	<b>v</b>		
Direction of displacement in shape deformation	<b>e</b>	$s''$	<b>d</b>			
Angle between shear or twinning direction and invariant line	$\alpha, \alpha$		$\alpha, \alpha$		$\alpha$	$\alpha$
Angle between shear or twinning plane normal and invariant normal	$\beta$		$\beta$		$\beta$	$\beta$
Angle turned through by shear or twinning plane in invariant line strain S	$\omega, \omega$		$\omega, \omega$		$\omega$	$\omega$
Angle turned through by macroscopic vector originally normal to habit plane	$\chi$			$\gamma$		

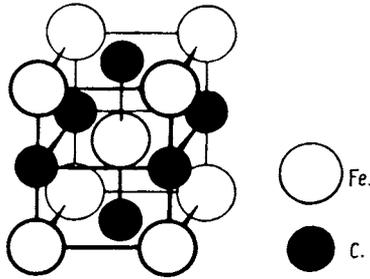


FIG. 22.8. The structure of steel martensites (after Petch, 1943). Only a small fraction of the carbon sites are occupied.

be denoted by  $\Lambda$  and  $M$ ; b.c.c. products will be regarded as merely a special case of b.c.t. phases.

As described in Section 89 martensite usually forms as lenticular plates but, in ferrous martensites, there is a wide variety of morphologies, many of which are not fully understood. These morphologies include  $\{557\}$  laths forming with carbon contents of less than 1.4%,  $\{225\}$  plates forming at intermediate carbon contents,  $\{259\}$  lenticular plates,  $\{3,10,15\}$  thin-plate martensite, h.c.p.  $\epsilon Z$  martensite in certain alloys of low stacking fault energy, "butterfly" martensite and surface martensite. With the exception of  $\epsilon Z$  martensite, the product in all cases is the same b.c.t. structure, but the internal structure varies considerably. In thin-plate martensite, fine twins extend right across the plate, in  $\{259\}$  and  $\{225\}$  martensites part of the cross-section is usually occupied by thin twins, but these may be confined to one interface (in  $\{225\}$  plates) or to the mid-rib region (in  $\{259\}$  plates), whereas the laths formed in low carbon alloys frequently contain numerous dislocations but no twins. Most of this section will be concerned with thin-plate and  $\{259\}$  martensites for which the crystallographic theory "works" well.

The observed orientation relations between the phases are rather similar for all alloys, a close-packed  $\{111\}_{\Lambda}$  plane of the austenite being approximately parallel to a  $\{011\}_M$  plane of the martensite in all cases. The relative directions of crystal axes within these planes vary from  $\langle 0\bar{1}1 \rangle_{\Lambda}$  approximately parallel to  $\langle 11\bar{1} \rangle_M$  (the Kurdjumov-Sachs orientation) to  $\langle 11\bar{2} \rangle_{\Lambda}$  approximately parallel to  $\langle 0\bar{1}1 \rangle_M$  (the Nishiyama-Wassermann relation). These two orientations differ by a rotation of about  $5\frac{1}{2}^{\circ}$  about  $\langle 111 \rangle_{\Lambda}$ ; the two most accurately determined orientations (the Greninger-Troiano relation for an iron-nickel-carbon steel and the relation in an iron-platinum alloy) being about midway between the two extremes.

The rather small range of orientations is associated with much larger variations in observed habit planes. For very low carbon steels (i.e. nearly pure iron), the data are rather confused, but the habit plane seems to be near  $\{111\}$  (Entwisle, 1956). For most carbon steels with less than 1.4% carbon, the Kurdjumov-Sachs orientation is found and the habit plane is  $\sim \{254\}_{\Lambda}$ . When the carbon content exceeds this value, there is a change in the habit plane (which becomes approximately  $\{259\}_{\Lambda}$ ) and the relative orientations approximate to the Nishiyama relation; this is also the situation in some iron-nickel

alloys. The accurately determined Greninger–Troiano orientation was associated with a habit plane  $\sim \{3,10,15\}$ . There are also a few experimental determinations of the direction of the shape deformation  $\mathbf{e}$  or rather of the component of  $\mathbf{e}$  resolved in the habit plane. The directions vary considerably, largely because of the experimental difficulties of measurement, but the angle  $\chi$  is about  $11^\circ$  in all cases.

The correspondence used in all theories of these transformations was first described by Bain (1924), who realized that a b.c.t. cell inscribed in the austenite could be changed to the correct dimensions of martensite by a contraction of about 17% along a cube axis and an expansion of about 12% in all directions at  $90^\circ$  to this axis. The compressive axis becomes the tetragonal  $c$  axis and the two  $\langle 110 \rangle$  axes normal to this become tetragonal  $\langle 100 \rangle$  directions.

A particular variant of the Bain strain<sup>†</sup> is given by the correspondence matrices

$$({}_M C_A) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad ({}_A C_M) = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & -1 & 0 \end{pmatrix} \quad (93.1)$$

so that the  $[010]$  cubic axis becomes the tetragonal  $c$  axis. Note that because two of the three principal deformations are equal to each other, there is only one unique principal axis,  $[010]$  in this case. Any two orthogonal directions in  $(010)$  may be used as the other principal axes, but it is most convenient to choose either  $[101]$  and  $[10\bar{1}]$ , or  $[100]$  and  $[001]$ . The principal deformations are  $\eta_1 = \eta_2 = \sqrt{2}a/a_0$ ,  $\eta_3 = c/a_0$ , where  $a_0$  is the length of the cube edge of the austenite unit cell, and  $c$  and  $a$  are the cell parameters of the martensite. In all alloys investigated,  $c/a$  varies from 1.0 to 1.08 and the ratio of the volumes of the two phases varies from  $\sim 1.03$  to  $\sim 1.05$ . The corresponding changes in  $\eta_1$  and  $\eta_2$  are quite small and do not greatly affect the predictions of the crystallographic theories. The Bain correspondence is illustrated in Fig. 22.9.

As both phases are single lattice structures, the correspondence relates all the positions of the iron atoms, and no shuffles are required. The carbon atoms, one of which is shown, occupy a fraction of the (equivalent) sites at the midpoints of the  $c$  edges or the centres of the faces normal to  $c$  in the martensite unit cell. These sites have distorted octahedral symmetry in relation to the iron atoms, and it might be thought at first that the choice of the austenite cube axis which becomes the martensite  $c$  axis (i.e. the choice of the carbon atoms in the correspondence) is determined by a local ordering of the carbon atoms in the austenite. Inspection of Fig. 22.9 shows that this cannot be so; the martensite structure is obtained automatically whichever axis becomes the  $c$  axis if the carbon

<sup>†</sup>This variant is chosen because it is used in most of the published work from Jaswon and Wheeler onwards. It should be noted that, although the same variant is used in the WLR paper on steels, the labelling of the martensite axes is different. Here standard crystallographic practice is used so that the unique  $c$  axis of martensite is labelled  $[001]_M$ ; in the WLR paper this axis is  $[010]_M$ . When other variants are considered, the results of this section may, of course, be applied by a suitable permutation of indices; austenite planes of the variant used by Bullough and Bilby, for example, are converted to the variant used here by the substitution  $h = h'$ ,  $k = l'$ ,  $l = -k'$ .

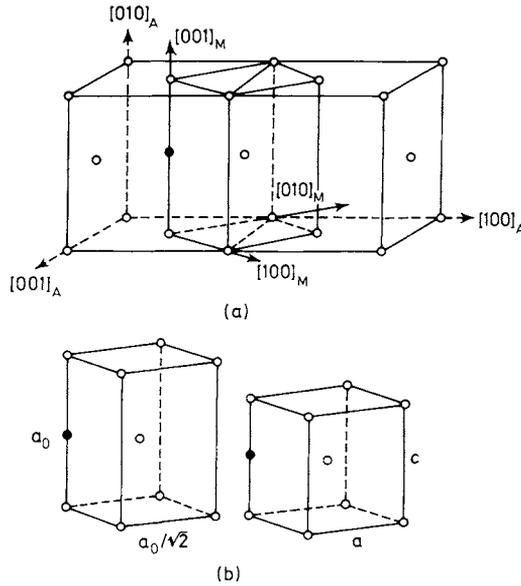


FIG. 22.9. The Bain correspondence: (a) b.c.t. cell of axial ratio  $\sqrt{2}$  outlined in the f.c.c. structure. The axes are labelled according to the variant of the correspondence given in eqn. (93.1). The filled circle indicates a carbon site. (b) The Bain strain carries the cell of axial ratio  $\sqrt{2}$  into the new cell of ratio  $c/a$ .

atoms are randomly distributed in the octahedral sites of the austenite prior to transformation. The transformation to martensite does not require any ordering of the carbon atoms although (paradoxically) the tetragonality may be ascribed to the preferential occupancy of one set of the interstitial sites of the b.c.c. structure. The relation of the carbon sites to the tetragonal unit cell is fixed by the particular variant of the correspondence which specifies the unique principal axis, and is presumably determined by the nucleation event.

The carbon sites in the austenite are symmetrically situated with respect to the three cube axes so that the way in which they are occupied cannot influence the selection of the unique axis.

The WLR analysis of the martensitic transformation assumed the product to be a stack of b.c.t. fine  $\{112\}_M$  twins with twinning direction implicitly assumed to be  $\langle 11\bar{1} \rangle_M$ , or else a single crystal with a lattice-invariant shear on the system  $\{112\}_M(11\bar{1})_M$ . The BM theory necessarily makes the same assumption, as the only mirror planes of the austenite lattice are  $\{100\}$  and  $\{110\}$ . From the correspondence [eqn. (93.1)] it follows that all three  $\{100\}$  planes become mirror planes of the martensite, and hence cannot act as twinning planes. Of the six  $\{110\}$  planes, those two which contain the  $c$  axis become mirror planes of the martensite, but the remaining four become crystallographically equivalent  $\{11\bar{2}\}_M$  planes. Choosing one of these,  $(110)_A$ , eqn. (92.45) shows that the direction  $\mathbf{l}$  is parallel to  $[1\bar{1}0]_A$ , and becomes  $\mathbf{l}^M$  parallel to  $\langle 11\bar{1} \rangle_M$ . Thus both the WLR and BM theories

make essentially the same assumption about the lattice-invariant shear, the elements of which are

$$(\mathbf{m}; \mathbf{A}) = (110) \quad [\mathbf{A}; \mathbf{I}] = [\bar{1}\bar{1}0] \tag{93.2}$$

The normal to the plane of shear is thus  $[001]_{\Lambda}$  and the habit planes are crystallographically degenerate (see discussion on p. 1009). Thus the four habit planes corresponding to this choice of the correspondence,  $\mathbf{m}$  and  $\mathbf{I}$  are all equivalent to each other, and there are two twin-related orientations which may be associated with each habit plane.

The crystallography of the steel transformations is particularly simple when the shear elements are those of eqn. (93.2); it corresponds to the two-dimensional case shown in Fig. 21.1. A parametric solution for the habit plane may readily be obtained using any of the procedures outlined in the previous section. Assuming the shape deformation to be an invariant plane strain, one variant of the habit plane is

$$[\Lambda; \mathbf{v}] = \frac{1}{2\eta_1(1 - \eta_2^2)^{1/2}} \begin{bmatrix} (\eta_1^2 + \eta_2^2 - 2\eta_1^2\eta_2^2)^{1/2} - (2 - \eta_1^2 - \eta_2^2)^{1/2} \\ (\eta_1^2 + \eta_2^2 - 2\eta_1^2\eta_2^2)^{1/2} + (2 - \eta_1^2 - \eta_2^2)^{1/2} \\ 2(\eta_1^2 - 1)^{1/2} \end{bmatrix} \tag{93.3}$$

The corresponding result for the BM assumption of a uniform dilatation in the interface is obtained by replacing  $\eta_i$  by  $\eta'_i = \delta\eta_i$ , as shown in the previous section. It is convenient to write  $\eta'_1 = \sqrt{2}\theta$  and  $\eta'_2 = \gamma\theta$ , where  $\theta = \delta(a/a_0)$  and  $\gamma$  is the axial ratio  $c/a$ . The habit plane indices are then

$$[\Lambda; \mathbf{v}] = \frac{1}{2\theta(1 - \gamma^2\theta^2)^{1/2}} \begin{bmatrix} \theta(1 + \frac{1}{2}\gamma^2 - 2\gamma^2\theta^2)^{1/2} - (1 - \theta^2 - \frac{1}{2}\gamma^2\theta^2)^{1/2} \\ \theta(1 + \frac{1}{2}\gamma^2 - 2\gamma^2\theta^2)^{1/2} + (1 - \theta^2 - \frac{1}{2}\gamma^2\theta^2)^{1/2} \\ 2(\theta^2 - \frac{1}{2})^{1/2} \end{bmatrix} \tag{93.4}$$

and substitution of  $\delta = 1$  gives the result (93.3) above. For a given axial ratio  $\gamma$ , the habit plane may be plotted as a function of a single variable  $\theta$ . The  $\mathbf{m}$  and  $\mathbf{I}$  restrictions impose limitations on the range of values of  $\theta$  (i.e. of  $\delta$ ) for which solutions are possible. These are

$$1/\sqrt{2} \leq \theta \leq [1 + \frac{1}{2}\gamma^2]^{1/2} \tag{93.5}$$

The extreme values of  $\gamma$  correspond to the limiting values of  $0^\circ$  and  $90^\circ$  respectively for the angle  $\alpha$  of the BM theory.

The habit plane (93.4) is one of the four variants which may be obtained from the correspondence (93.1) and shear elements (93.2). [It is the  $(\alpha_+, \omega_+)$  solution of the BM theory.] As already described, the variants corresponding to different values of the lattice-invariant shear (or to different signs for  $\omega'$ ) are related by reflections in the plane  $\mathbf{m}$  and those corresponding to different choices of the invariant line (different  $\alpha$ 's) are related by

rotation of  $180^\circ$  about the normal to the plane of shear  $[001]_\Lambda$ . Thus the four solutions may be written

$$\begin{aligned}(\alpha_+ \omega_+)(\mathbf{v}; \mathbf{A}) &= (hkl) \\(\alpha_+ \omega_-)(\mathbf{v}; \mathbf{A}) &= (\bar{k} \bar{h}l) \\(\alpha_- \omega_+)(\mathbf{v}; \mathbf{A}) &= (\bar{h} \bar{k}l) \\(\alpha_- \omega_-)(\mathbf{v}; \mathbf{A}) &= (khl)\end{aligned}\tag{93.6}$$

The relation between the four solutions is shown in Fig. 22.10, where the habit plane is plotted as a function of  $\theta$  for b.c.c. products ( $\gamma = 1$ ). The solutions lie in four quadrants of the stereographic projection, the other four quadrants corresponding to the solutions using  $\mathbf{m} = (1\bar{1}0)$ ,  $\mathbf{l} = [110]$ .

In Fig. 22.11, the habit plane of eqn. (93.4) is plotted as a function of  $\theta$  for the axial ratios 1.00 and 1.08. Following Wechsler *et al.* (1953), the habit plane is also plotted for fixed volume ratios of 1.03 and 1.05. Thus, in the absence of dilatation, the theoretical result is that all habit planes should lie in the small area  $ABCD$  of the stereographic projection of Fig. 22.11. The predicted habit plane is obviously rather insensitive to the composition of the alloy, that is, to the lattice spacings of the two phases.

Figure 22.11 shows the experimental habit plane of Greninger and Troiano (1949) for an iron-22.5% nickel 0.8% carbon alloy, together with the theoretical prediction for this alloy using the parameters  $\gamma = 1.045$  and  $\theta^2 = (a/a_0)^2 = 0.627$  given by the authors.

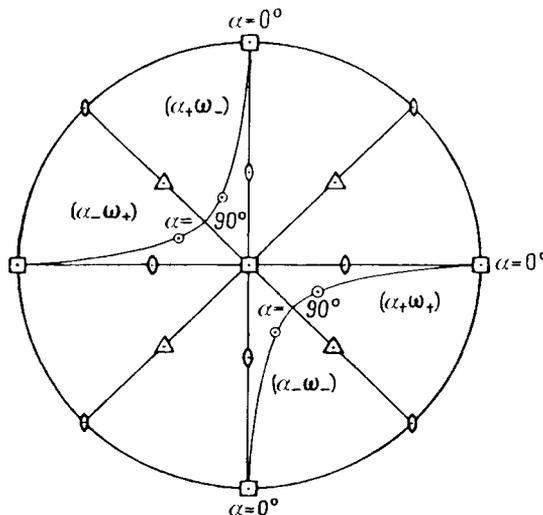


FIG. 22.10. Stereographic projection showing the variation of the habit plane with  $\theta$  for b.c.c. martensite ( $\gamma = 1$ ) (after Bowles and Mackenzie, 1954b). As  $\theta^2$  varies between the limits  $1/2$  and  $2/3$ ,  $\alpha$  varies from  $0^\circ$  to  $90^\circ$ . The four circles represent the solutions obtained for the habit planes for some particular value of  $\theta$ .

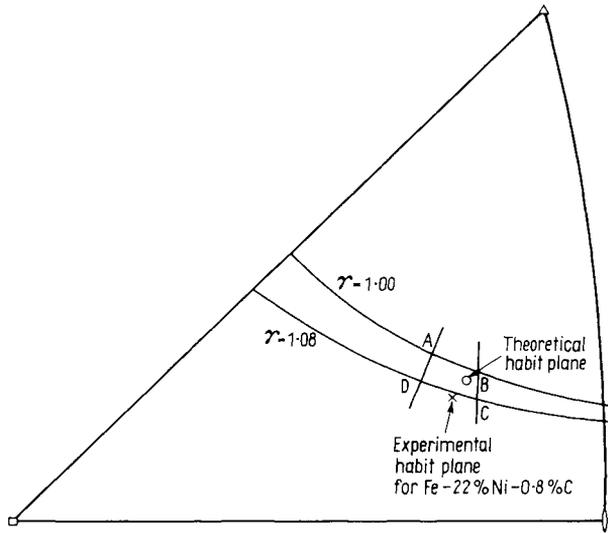


FIG. 22.11. Predicted habit planes for steels. According to the Bowles Mackenzie theory, the observed habit plane for a particular transformation should lie somewhere along the curve for the appropriate axial ratio of the martensite produced. The curves for  $\gamma = 1.00$  and  $\gamma = 1.08$  represent the extremes normally encountered. Curves of constant volume ratio 1.03 and 1.05, assuming  $\delta = 1$ , are also plotted; in the absence of dilatation, all habit planes should lie inside the area  $ABCD$ . The figure also shows the habit plane measured accurately by Greninger and Troiano (1949) and the corresponding theoretical prediction, assuming  $\delta = 1$ .

There is excellent agreement between theory and experiment without the need for any dilatation in the interface. Bowles and Mackenzie (1954b) used the value  $\theta^2 = 0.630$  (corresponding to  $\delta = 1.002$ ) but it is not clear that this gives any better agreement with experiment than the assumption that  $\delta = 1$  (see Fig. 22.11).

Figure 22.12 shows the results of experimental habit plane determinations in two iron-nickel alloys. There is an alarming scatter in the experimentally observed habit plane in each alloy.

The specification of the shape deformation is now completed by the unit vector  $\mathbf{e}$ , which may also be readily expressed in terms of  $\eta_i$  or of  $\theta$  and  $\gamma$ , and the magnitude of the angle  $\chi$  given by

$$\tan \chi = (2\theta^2 - 1)^{1/2} (1 - 2\gamma^2\theta^2)^{1/2} / 2\gamma^2\theta^2 \tag{93.7}$$

The angle  $\chi_0$  is correspondingly given by

$$\tan \chi_0 = (2\theta^2 - 1)^{1/2} (1 - 2\gamma^2\theta^2)^{1/2}$$

Greninger and Troiano treated the shape deformations as a simple shear and measured the approximate shear angle  $\chi'$ , which is related to the angle  $\chi_0$  by the equation  $2\tan(\chi'/2) = \tan\chi_0$ . The mean experimental value of  $\chi' = 10.75^\circ$  corresponds to  $\chi_0 = 10.66^\circ$  and agrees very well with the theoretical predictions  $\chi = 10.33^\circ$  and  $\chi_0 = 10.61^\circ$  which are

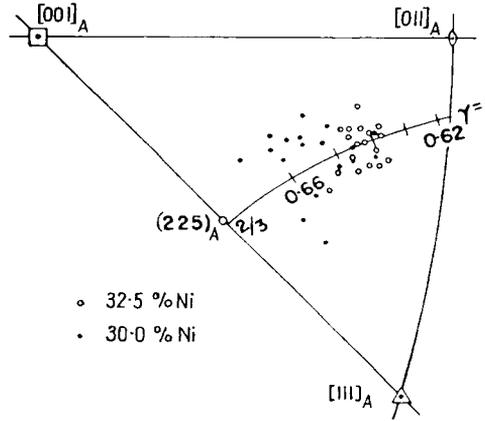


FIG. 22.12. Comparison of the predicted and measured habit planes for transformations with  $\gamma = 1$  (after Bowles and Mackenzie, 1954b). The results for 30.0% Ni and 32.5% Ni are those obtained respectively by Machlin and Cohen (1951) and by Greninger and Troiano (1949).

obtained for  $\delta = 1$ ; the actual measurements scatter from about  $8^\circ$  to  $14^\circ$ . The scatter in the observed direction of shear is very large, about  $30^\circ$ , so that comparison with theory is not possible. However, the mean direction is close to the theoretical prediction and Bowles and Mackenzie gave an analysis to show that the experimental scatter is to be expected because of the assumption made by Greninger and Troiano that the shape deformation is a simple shear.

It remains only to compare the orientation relations which cannot be conveniently expressed in parametric form. For any habit plane, two twin-related orientations are predicted, and the experimental observation that some plates contain large twins shows that both orientations may be present together. More recently, electron microscopy has given direct evidence of fine twinning in some plates (see p. 1054). Greninger and Troiano, however, identified only one major orientation and, on comparison with the theory, it becomes evident that this orientation corresponds to the smaller lattice-invariant strain. This orientation is the BM ( $\alpha_1 \omega_1$ ) solution; the other orientation is given by the ( $\alpha_1 \omega_-$ ) solution. The fraction of the twinning shear is given by

$$f^s = 1/2 \pm 1/2 [1 - 2(2 + \gamma^2)(1 - \gamma^2 \theta^2)(2\theta^2 - 1)\theta^{-3}(2 - \gamma^2)^{-2}]^{1/2} \quad (93.8)$$

and substitution of the parameters gives  $\sim 0.4$  and  $0.6$  for the two values of  $f^s$ .

The relative orientations of the austenite and martensite given by the  $\alpha_1 \omega_\pm$  solution are compared with the experimental values in Table XXII taken from the BM paper. This table gives the angles between corresponding directions and planes in the two structures.

Once again, the agreement is very good. As already stated, this orientation relation is intermediate between that of Kurdjumov-Sachs, for which the first pair of directions in the table are parallel, and that of Nishiyama, for which the third pair of corresponding directions are parallel.

TABLE XXII. COMPARISON OF PREDICTED AND MEASURED ORIENTATION RELATIONS in Fe-22% Ni-0.8% C

Corresponding directions	Measured angle	Predicted angle
$[\bar{1}\bar{1}0]_A - [11\bar{1}]_M$	2.5°	2.7°
$[0\bar{1}1]_A - [11\bar{1}]_M$	6.5°	6.6°
$[1\bar{2}1]_A - [10\bar{1}]_M$	2.0°	1.9°
$(111)_A - (101)_M$	~1°	0.2°

More recent work has shown that two types of morphology are involved in the above characteristic crystallography of martensite in steels containing amounts of carbon in excess of 1.4% and with  $M_s$  temperatures below room temperature. The best agreement with the standard theory is obtained from so-called "thin-plate martensite" in which the habit plane interfaces are very straight and the twin substructure is uniform and extends right across the plate from one interface to another (see Fig. 22.13). In other alloys, martensite plates are lenticular in shape and have a much less regular internal structure, sometimes consisting of a heavily twinned layer confined to the vicinity of the central plane or "mid-rib" of the plate (see Fig. 22.14). Such plates are generally referred to as {259} but, as shown in Fig. 22.12, they have a wide scatter of habits to be reconciled with the crystallographic theories. It is generally supposed that in the nucleation and early growth stages, the lattice-invariant shear is produced by twinning, but that as growth proceeds the released latent heat warms up the plate into a higher temperature region at which growth by a dislocation interface is preferred. No mechanism for the transition from twinned to untwinned regions of the plate has been suggested.

McDougall and Wayman (1992) have summarized the most reliable data on the {3,10,15} and {259} types of transformation as those for Fe-Ni-C (Greninger and Troiano, 1949; Dunne and Bowles, 1969), Fe-Al-C (Watanabe and Wayman, 1970ab) and Fe-Pt (Efsic and Wayman, 1967). The experimental and theoretical values, as compiled by Wayman, are shown in Table XXIII and the agreement is impressive. The value of  $\delta$  is effectively unity in all cases.

Any complacency about the theory's success in this transformation is soon removed when the results of other determinations of the habit plane are considered. There are two main difficulties. Firstly, many habit planes are well outside the small region near {3,10,15} which the above theory predicts for all alloys and, secondly, individual determinations of the habit plane for any one alloy often show a wide scatter. An indication of this is given in Fig. 22.12, which shows the predicted habit planes for  $\gamma=1$  and various values of  $\theta^2$  together with two independent sets of experimental results for iron-nickel alloys containing 30% and 32% nickel respectively.

Later work (Otte and Read, 1957) has confirmed that the experimental scatter is a real effect, and is not due to experimental difficulties inherent in identifying the true habit of

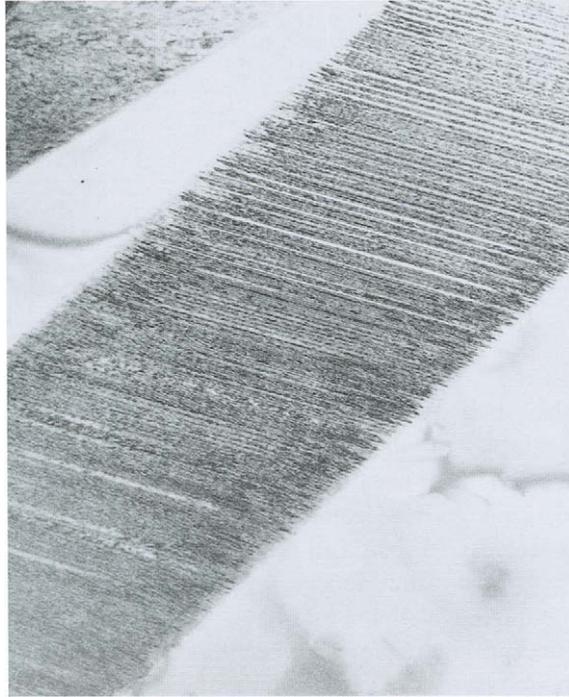


FIG. 22.13. Finely twinned microstructure of  $\{3, 10, 15\}$  thin-plate martensite to illustrate how twins fill the whole cross-section of the plate.

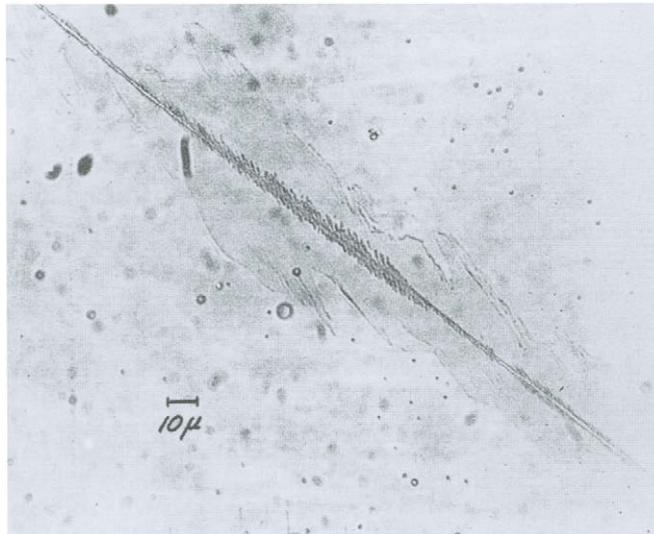


FIG. 22.14. Twins along the mid-rib of a  $\{254\}$  plate in an iron nickel alloy (courtesy of Patterson and Wayman, 1966).

TABLE XXIII. COMPARISON BETWEEN PREDICTED AND EXPERIMENTAL VALUES FOR THE (3, 10, 15) TRANSFORMATION

	Fe-22 Ni-0.8C		Fe-7 Al-2C		Fe-24.5Pt(at.%)	
	Exp. <sup>1,3</sup>	Calc. <sup>4</sup>	Exp. <sup>4</sup>	Calc. <sup>4</sup>	Exp. <sup>2</sup>	Calc. <sup>2</sup>
(111) <sub>F</sub> : (101) <sub>B</sub>	1.0	0.2	0.0	0.1	0.86	0.68
[111] <sub>F</sub> : [101] <sub>B</sub>	2.5	2.7	2.67	2.55	0.68	4.06
(P <sub>1</sub> ) <sub>F</sub> :	0.1748	0.1694	0.1267	0.1227	0.1910	0.1693
	0.5481	0.5964	0.5756	0.6045	0.6214	0.7613
	0.8180	0.7857	0.8078	0.7874	0.7599	0.6259
[d] <sub>F</sub> :	-0.1761	-0.1878	-0.1398	-0.1325	-0.1275	-0.1907
	-0.6886	-0.6657	-0.6303	-0.6545	-0.7292	-0.7050
	0.7034	0.7222	0.7636	0.7444	0.6723	0.6831
m <sub>1</sub>	0.1915	0.1988	0.1220	0.1379	0.2157	0.2292
m <sub>2</sub>	...	0.2294	...	0.1622	...	0.2448
δ	1.002	1.000	1.002	1.000	0.997	1.000

**Key to references**

- 1 Greninger and Troiano (1949)
- 2 Efsic and Wayman (1967)
- 3 Dunne and Bowles (1969)
- 4 Watanabe and Wayman (1971)

a lenticular plate. The scatter is reduced but not eliminated when the habit is defined by the mid-rib rather than the edge of the plate (Breedis and Wayman, 1962). It seems most likely that a difference in habit between one plate and another, even in the same austenite grain, is caused by the coherency stresses which oppose the development of each martensite plate. The invariant plane strain condition, as stated in Section 89, is strictly applicable only to an unconstrained transformation with the interface traversing a single crystal. The formal effect of the constraints might be to modify the invariant plane strain condition or to imply a more complex lattice-invariant deformation. The scatter across the lines of constant  $\gamma$  is as large as the scatter along these lines so that the simplest modification (a variable but uniform dilatation in the habit plane) does not explain the results. Nevertheless, Bowles and Mackenzie have shown that the mean habit plane of those plotted in Fig. 22.12, as well as the data of others on iron-carbon alloys with {259} habits, can be explained reasonably well with a dilatation parameter which is about  $\delta = 1.004$  for most alloys. They also gave a detailed analysis of the shape deformation measured by Machlin and Cohen (1951); the experimental estimates were that the shear and normal components of the shape change were 0.20 and 0.05 and the corresponding theoretical values with  $\delta = 1.004$  were 0.219 and 0.041.

Turning from the scatter of habit planes at one composition to the problem of habits not predicted by the above theory, there are two main proposals to consider. One of these is the BM assumption of a dilatation parameter; the other is the possibility that the lattice-invariant deformation does not utilize the twinning elements of the product phase assumed above. In medium carbon steels, the habit plane is near to a {225} austenite plane and this

has caused great difficulties in interpretation. In the model developed by Frank (1953) the close-packed planes of the two structures meet edge to edge in the interface along a common close-packed direction. Thus the invariant line  $\mathbf{d}$  becomes identical with the twinning direction  $[1\bar{1}0]_A$  or  $[11\bar{1}]_M$ . The difference in the interplanar spacing is compensated by a slight mutual tilt of the two planes, but the difference in repeat distance along the  $[1\bar{1}0]=[11\bar{1}]$  axis has to be accommodated as elastic strain.

With the exception of thin-plate martensite, the internal structure is usually very non-uniform and it is difficult to devise a growth model that will include these heterogeneities. Most  $\{259\}$  and  $\{225\}$  plates contain some  $\{112\}$  twins, and this suggests that such twinning makes a substantial contribution to the lattice-invariant deformation. Many attempts have been made to modify the above theory by utilizing a different plane or direction of the lattice-invariant shear, or by using two lattice-invariant shears.

There are many difficulties associated with Frank's simple model. The discrepancy in the lattice spacings  $d_{101}$  and  $d_{111}$  is about 2% in various steels and the small angle  $\psi$  between the two almost parallel planes is about  $\frac{1}{2}^\circ$ , so that the relation deviates very little from the ideal Kurdjumov-Sachs orientation. Another way of matching the structures is to use the BM dilatation parameter. From Fig. 22.10, it will be seen that this requires a maximum value of the parameter  $\alpha$ , i.e.  $90^\circ$ , and the invariant line becomes identical with the twinning direction  $[1\bar{1}0]_A$  or  $[11\bar{1}]_M$ . With available lattice parameters, the value of  $\delta$  to give  $\alpha = 90^\circ$  is 1.013-1.015 and  $\theta^2$  is given the maximum value of  $(1 + \frac{1}{2}\gamma^2)^{-1/2}$ . For  $\gamma = 1$ ,  $\omega = 0$  and  $\omega$  never exceeds  $20'$  over the whole range of  $\gamma$ , in agreement with the experimental result that 225 plates have nearly Kurdjumov-Sachs orientation relations. If  $\alpha = 90^\circ$ , both the  $(\alpha, \omega_+)$  and the  $(\alpha, \omega_-)$  solutions represent the same strain  $\mathbf{S}$  and the orientation relations of the two twins with the matrix are equivalent variants of each other. The shear  $k^M$  is exactly one-half of the twinning shear and the habit plane is given by

$$(\mathbf{v} : \mathbf{a}) = \frac{1}{2\sqrt{2}} \left\{ (2 - \gamma^2)^{1/2}, (2 + \gamma^2)^{1/2}, (4 + 2\gamma^2)^{1/2} \right\}$$

For  $\gamma = 1$ , the predicted habit plane is  $\{2, 2, 4.90\}_A$ .

The variation of the angles  $\phi$  and  $\psi$  is shown for the BM theory and Frank's model in Fig. 22.15, which was used by Bowles and Mackenzie in a comparison of the two theories. The matching of the two structures along the close-packed direction, where there is a discrepancy of almost 2%, is accomplished by a dilatation parameter in the BM theory and by an anisotropic misfit related to an assumed Poisson's ratio in the Frank model. The two assumptions have opposite effects on the angles  $\phi$  and  $\psi$  as the axial ratio increases from unity. Early work on an iron - 7.9% chromium - 1.11% carbon alloy by Wayman *et al.* (1961) apparently gave better agreement with Frank's model.

However, neither description is able to account for all the detailed experimental results. Moreover, both descriptions require that the habit plane differs from an exactly invariant plane and this assumption has been directly tested by taking replicas from the surface

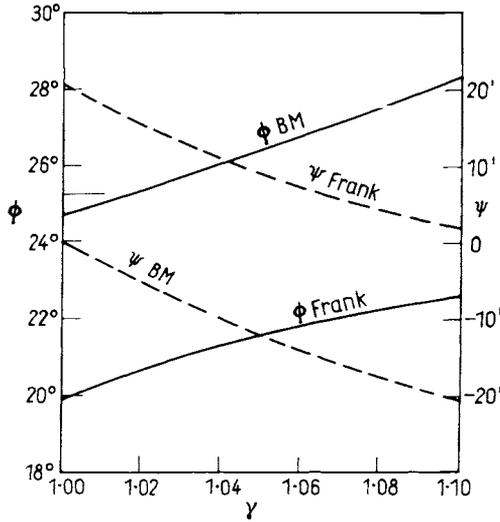


FIG. 22.15. Comparison of the BM theory and Frank's theory for {225} martensite (after Bowles and Mackenzie, 1954b).  $\psi$  is the angle between  $(111)_A$  and  $(101)_M$  and  $\phi$  is the angle between the habit plane and  $(111)_A$ .

(Krauklis and Bowles, 1969; Muddle *et al.*, 1976). These experiments showed that changes in length in the interface were certainly less than 0.2% and are probably zero.

The possibility that the lattice-invariant deformation employed in some cases slip planes and directions other than those corresponding to twinning elements of the martensitic structure was investigated very thoroughly by Crocker and Bilby (1961). They utilized all possible known slip and twinning modes in both austenite and martensite and worked with lattice parameters appropriate to Greninger and Troiano's alloy, after first checking that the predicted habit planes were not sensitive to the lattice parameter in any mode. The results for all the non-equivalent variants of the major deformation modes are shown in Table XXIV, which gives the habit plane indices, the magnitude of the shape deformation and the magnitude of the lattice-invariant shear. The habit plane results are also plotted in Fig. 22.16, which also shows the total scatter regions for various sets of habit plane data. Only three of the 20 predictions fall within any of the scatter regions and no habit plane is predicted within the scatter regions of the observed {225} and {111} habits.

The surface dislocation theory in its original form supposed that  $\mathbf{l}$  is a rational direction (being a lattice Burgers vector) and that  $\mathbf{m}$  is probably also rational (being a slip plane). Actually, it will be shown in the next section that the BM twinning hypothesis requires non-rational  $\mathbf{l}$  in some cases, but it seems at first natural to suppose that  $\mathbf{l}$  is rational if the lattice-invariant shear is accomplished by dislocations in the interface. However, a surface array of parallel dislocation lines could be the effective resultant of a number of arrays of parallel lines provided all the lines have the same Burgers vector or lie in the same slip plane. In the latter case, the effective Burgers vector may be irrational. Similar

TABLE XXIV. MAJOR DEFORMATION MODES FOR STEELS INVESTIGATED BY CROCKER AND BILBY

Number		Deformation mode ( $\mathbf{m}$ or $\mathbf{m}^M$ and $\mathbf{l}$ or $\mathbf{l}^M$ )		Shear (g)	Shape Shear ( $\chi$ )	Habit Planes	
Austenite modes	1	$(\bar{1}\bar{1}\bar{1})$	$[\bar{1}0\bar{1}]$	0.36	27°	{239}	{133}
	2	$(\bar{1}\bar{1}\bar{1})$	$[\bar{1}\bar{1}\bar{0}]$	0.23	13°	{267}	{235}
	3	$(\bar{1}\bar{1}\bar{1})$	$[\bar{2}\bar{1}\bar{1}]$	0.24	18°	{011}	{237}
	4	$(\bar{1}\bar{1}\bar{1})$	$[\bar{1}\bar{2}\bar{1}]$	0.26	10°		{456}
Martensite modes	5	$(\bar{1}\bar{1}\bar{0})$	$[111]$	0.43	28°	{689}	{112}
	6	$(\bar{1}0\bar{1})$	$[111]$	0.23	13°	{267}	{235}
	7	$(\bar{2}\bar{1}\bar{1})$	$[111]$	0.27	20°	{349}	{568}
	8	$11\bar{2}$	$[111]$	0.23	10°		{3, 10, 15}
	9	$1\bar{2}\bar{3}$	$[111]$	0.23	11°	{358}	{067}
	10	$\bar{3}1\bar{2}$	$[111]$	0.25	17°	{349}	{234}
	11	$\bar{2}\bar{3}\bar{1}$	$[111]$	0.31	22°	{112}	{223}

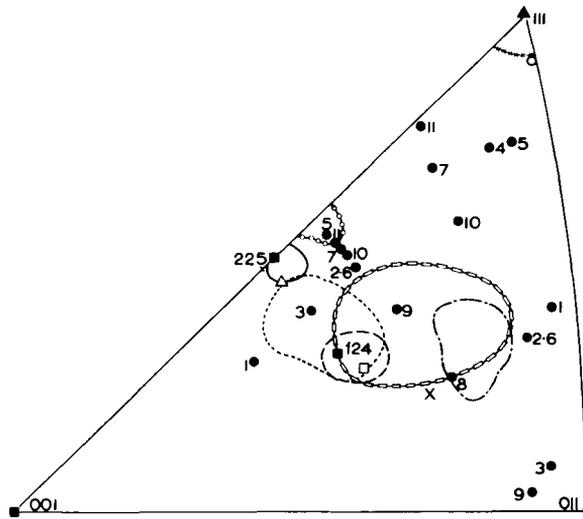


FIG. 22.16. Experimental and predicted habit planes for martensites in steels (after Crocker and Bilby, 1961). Regions of scatter of experimental habit plane results: (solid line) Fe 0.95% C, Fe 1.4% C (Greninger and Troiano, 1940); (long dashes) Fe 1.78% C (Greninger and Troiano, 1940); (alternate long/short dashes) Fe 32.5% Ni (Greninger and Troiano, 1940); (open rectangles) Fe 30.0% Ni (Machlin and Cohen, 1940); (short dashes) Fe 2.8% Cr 1.5% C (Otte and Read, 1957); (crosses) pure Fe (Entwisle, 1956); (open circles) Fe 7.9% Cr 1.11% C (Wayman *et al.*, 1961); (single cross) Fe 22% Ni 0.8% C (Greninger and Troiano, 1949). Predicted habit planes: (●) results for possible modes of mechanical deformation listed in Table XXIV; (○)  $\{111\}_A$  habit predicted by lattice-invariant shear on  $(\bar{1}\bar{1}\bar{0})_M$   $[110]_M$  or by equal amounts of slip on  $(\bar{1}\bar{1}\bar{0})_M$  in the  $[111]_M$  and  $[\bar{1}\bar{1}\bar{1}]_M$  directions; (□)  $\{259\}_A$  habit predicted by shear on  $(\bar{3}\bar{2}\bar{4})_A$   $[2\bar{1}\bar{1}]_A$ ; (△)  $\{225\}_A$  habit predicted by shear on  $(\bar{1}\bar{1}, 1\bar{2}, \bar{1}\bar{0})_A$   $[211]_A$ .

considerations apply to the plane of the lattice-invariant deformation, so it may be necessary to examine the possibility that only one of the two elements  $\mathbf{l}$  and  $\mathbf{m}$  is rational.

Crocker and Bilby made a detailed study of the effects of this more general assumption. They used the following low index planes and directions as fixed elements of the lattice-invariant deformation

$$\text{Directions: } \langle 100 \rangle_{\Lambda} \langle 110 \rangle_{\Lambda} \langle 112 \rangle_{\Lambda} \langle 100 \rangle_{\text{M}} \langle 110 \rangle_{\text{M}} \langle 111 \rangle_{\text{M}}$$

$$\text{Planes: } \{100\}_{\Lambda} \{011\}_{\Lambda} \{111\}_{\Lambda} \{100\}_{\text{M}} \{011\}_{\text{M}} \{112\}_{\text{M}} \{123\}_{\text{M}}$$

These include all the slips and twinning elements of the two phases and there are 15 non-equivalent variants for each correspondence of which two, (010) and  $[010]_{\Lambda}$ , gave no solution. For the remaining 13 fixed elements, the habit planes were calculated at  $5^{\circ}$  intervals as the variable element rotated through  $90^{\circ}$  and  $180^{\circ}$  depending on the symmetry.

The deformation modes studied were thus divided into 13 sets, each with either 19 or 37 variable elements. For each deformation mode results were calculated with the Greninger and Troiano parameters and also with a dilatation of  $S = \pm \frac{1}{2}\%$ . In some cases the  $\mathbf{m}$  or  $\mathbf{l}$  restrictions became operative, but over 3000 habit planes were calculated with the assistance of an early electronic computer.

The very large number of theoretical predictions now made it difficult to compare with experiment. Crocker and Bilby argued, however, that for any fixed element the variable element should probably be selected by minimizing either the magnitude of the lattice-invariant shear or that of the total shape distortion. Thus they calculated the values of  $k$  and of  $\chi$  with varying  $\mathbf{m}$  or  $\mathbf{l}$  respectively for each system and considered only the results corresponding to minimum shears. Each system had a minimum for both shears, and in five of the systems the two minima coincided and were the absolute minima over all systems. In these five systems neither  $\mathbf{m}$  nor  $\mathbf{l}$  restrictions are operative and habit plane results were obtained over the whole range of variation of the rotational element.

For several systems, the minimum shear criterion indicates that the most probable modes for the lattice-invariant shear are rational and unique, corresponding to coincident minima for both  $k$  and  $\chi$ . Most of the modes indicated in this way corresponded to usual deformation modes included in Table XXIV. In addition, double shear modes on  $\{123\}$  planes that contain only one close-packed direction are unlikely.

This leaves the modes arising from the following fixed elements for further consideration:  $-[100]_{\Lambda}$ ,  $[211]_{\Lambda}$ ,  $[110]_{\text{M}}$  and  $[121]_{\text{M}}$ . The minima of  $k$  and  $\chi$  coincide only for  $[101]_{\Lambda}$  and  $[110]_{\text{M}}$ ; for the others, the minimum  $k$  results were evaluated. These four systems give 10 habit planes of which one is  $3^{\circ}$  from  $\{111\}_{\Lambda}$  and another is near  $\{259\}_{\Lambda}$ . These results are also shown in Fig. 22.16.

A habit plane near  $\{111\}_{\Lambda}$  is given by the lattice-invariant shear  $\{110\} [110]_{\text{M}}$  and this is shown by an open circle in Fig. 22.16. Although this is not a normal deformation mode and so is not included in Table XXIV it may be regarded as equivalent to equal amounts of shear on the two equivalent  $\langle 111 \rangle$  directions in  $\{011\}$ . From the magnitude of the slip

shear it seems reasonable to conclude that if slip is to occur on a  $\{011\}$  plane it will prefer to utilize both  $\langle 111 \rangle$  directions, thus giving a habit near  $\{111\}$ .

Now consider the so-called  $\{259\}$  habits, the main problem with which is the lack of consistency. None of the rational modes in Table XXII predict these habits but a more general shear on  $\langle 321 \rangle \{211\}$  gives a habit close to  $\{259\}$ . This slip system could essentially be a combination of slip on a  $\{111\}$  plane in a  $\langle 211 \rangle$  direction and a small amount of  $\{011\} \langle 211 \rangle$  slip. Varying proportions of the two systems will then account for the scatter as the habit plane varies between the minimum  $k$  and the minimum  $\chi$  situation. Scatter at right angles to this might be accounted for by mixing modes of 1 and 2, 6, and indeed mode 3 may be regarded as an equal mixture of these two modes. The experimental results might thus be explained if both the plane and direction of the lattice-invariant shear are varied to give the necessary two parameters. However, there are difficulties in understanding how an irrational plane and direction could lead to a glissile interface.

Since this early work, a number of attempts to produce a self-consistent theory of  $\{225\}$  martensite have all concluded that it is impossible to do this simply by varying the assumptions of the formal crystallographic theory. There is general agreement that the habit is actually close to  $\{449\}$  but that it is not of form  $\{h h k\}$ . Most treatments have assumed that the observed twinning systems constitute a major part of the lattice-invariant deformation because dense twins are found to occur on one very straight interface of the plates. A careful analysis by Wayman and Muddle (1970) shows that it is seemingly impossible to find a description which will account for both the habit plane and the shape deformation observations. Bowles and Dunne (1961) developed a plastic accommodation theory in which deformation in the matrix ahead of the plate compensates for the discrepancy between the interatomic distances along the  $[\bar{1}10]_A$  and  $[111]_B$  planes which are assumed to be parallel in the interface. This model accounts for many of the features of the  $\{225\}$  transformation but not for others. A proposed double shear form of the theory (Ahlers, 1974, 1975) in which the shape deformation is written  $\mathbf{E} = \mathbf{R} \mathbf{P}_2 \mathbf{G}_2 \mathbf{P}_1 \mathbf{G}_1$ , where  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are homogeneous pure lattice deformations and  $\mathbf{K}_1$  and  $\mathbf{K}_2$  are shears on  $(001)[\bar{1}\bar{1}1]_B$  and a twinning shear on  $(1\bar{1}2)[\bar{1}11]_B$  respectively, is stated rather surprisingly to give results almost identical to those of Wechsler *et al.* (1953). However, this has the wrong twinning plane  $(\bar{1}\bar{1}2)_B$  instead of the observed  $(112)_B$ , and the physical significance of dividing the pure lattice deformation  $\mathbf{P}$  into two parts is also not clear. Finally, an analysis by Sandvik and Wayman (1983) for a subplate utilizes a single lattice-invariant shear on  $(11\bar{3})_A$  in the direction  $[\bar{1}10]_A$  which is based on an estimate of the invariant line. The solutions are in good agreement with experiment apart from the magnitude and direction of the shape strain. The calculated invariant line is very close to a set of dislocations found in the interface (see p. 1030). The lattice-invariant deformation of the macroscopic plate is assumed to differ from that of the sub-plates so that the shape deformation  $\mathbf{E} = \mathbf{R} \mathbf{P} \mathbf{K} \mathbf{K}_2$  and the rotation  $\mathbf{R}$  and shape strain  $\mathbf{E}$  are specified from experiment.

Kelly (1992) has proposed a double shear theory of  $\{225\}$  martensite which is claimed to account for all the experimental features. He supposes the main shear system to have

at least a component of the twinning system  $(112)_M[1\bar{1}\bar{1}]_M$ . The shear direction is taken to be  $[1\bar{1}\bar{1}]$  but the shear plane is allowed to be any plane between  $(112)_M$  and  $(\bar{1}10)_M$ , i.e. a combination of the two rational systems. Trial and error then shows that the  $K_2$  shear system would have to be irrational, but the final choice was  $(122)_A[\bar{2}\bar{2}\bar{1}]_A$ . A small amount of shear on this system combined with the  $K_1$  shear gives a habit plane very close to the observed plane as the shear plane of the  $K_1$  system was varied from  $(44\bar{1})_A$  to  $(11\bar{3})_A$ , and the orientation relation is also exactly predicted. For the shape deformation, however, the same variation in the plane of the first lattice-invariant shear resulted in a spread of over  $50^\circ$  in the direction, and Kelly emphasizes that this is also a feature of the experimental results for  $\{225\}$  martensite, as distinct from  $\{259\}$  or  $\{3,10,15\}$  martensites which have small spreads of about  $6^\circ$ . Whether or not this rather elaborate theory gives a true description of  $\{225\}$  martensite remains to be seen.

Martensite in low carbon alloys is generally described as lath martensite and is assigned a habit plane near  $\{111\}_A$  or, more accurately,  $\{557\}_A$ . This martensite is not twinned but contains an appreciable density of all four f.c.c. screw dislocations (see Fig. 22.17). Adjacent laths are arranged in parallel "packets" and give a "blocky" type structure which is rather similar to that of bainite. A complication is that, because this type of martensite forms at an  $M_s$  temperature of about  $250^\circ\text{C}$ , specimens are normally completely transformed without any retained austenite when examined at room temperature, and the martensite is probably auto-tempered to some unknown extent. This may be avoided by alloying, as certain alloys (e.g. Fe-Ni-Mn) produce lath-like structures even though their  $M_s$  temperatures are slightly below room temperature.



FIG. 22.17. Lath martensite in an iron-nickel-carbon alloy (courtesy of Kelly and Nutting, 1960 1).

Adjacent laths have generally been reported to have almost the same orientation but also, in carbon-free materials, to have misorientations which accumulate in a given sense (Rao and Thomas, 1979). Twin-related laths have been widely reported in carbon-containing materials. The orientation relationship with the austenite has been reported as the Kurdjumov–Sachs, the Nishiyama–Wassermann and various intermediate orientations.

There is at present no successful theory of the crystallography of lath martensite. Kelly *et al.* (1982) supposed two lattice-invariant shears having the shear plane normals and shear directions in a common plane, this being equivalent to a single shear plane plus a rigid body rotation. The shear components had magnitude  $g_1$  on  $(\bar{1}\bar{1}3)[\bar{1}\bar{1}0]_A = (12\bar{1})[11\bar{1}]_B$  and  $g_2$  on  $(\bar{1}11)[0\bar{1}1]_A = (\bar{1}\bar{2}1)[1\bar{1}\bar{1}]_B$  with  $g_2 > g_1$ . For  $g_2 = -0.6$ , the locus of the habit plane normal crosses the  $(\bar{1}\bar{1}0)_A$  zone at a position that is several degrees from the observed (557) pole, but is within the range of values for the fully transformed structures.

Sandvik and Wayman (1988) have proposed another approach, stemming from the observation of what they consider to be misfit dislocations in the interface. They consider the line direction of these dislocations to be the invariant line and deduce that this represents the meeting of the  $(100)_A$  and  $(211)_B$  planes. These planes are not corresponding planes of the Bain strain and the direction of the interface dislocations is not an undistorted direction. Instead, Sandvik and Wayman propose a new correspondence

$${}^{(B)C}_A = \frac{1}{2} \begin{pmatrix} 1 & \bar{1} & 1 \\ 1 & \bar{1} & \bar{3} \\ 1 & \bar{3} & 1 \end{pmatrix}$$

The calculations were carried out in the form of double shear equations and the solutions are in good agreement with experiment with the exception of the magnitude of the shape deformation, for which the predicted value is 0.96 and is very much larger than the observed values in the region of 0.3.

Kelly (1992b) has given a two shear analysis of lath martensite somewhat analogous to his treatment of  $\{225\}$  martensite. The first lattice-invariant shear is assumed to be in the direction  $[11\bar{1}]$  on any plane between  $(\bar{1}10)_M$  [i.e.  $(00\bar{1})_A$ ] and  $(112)_M$  [i.e.  $(110)_A$ ]. The second lattice-invariant shear  $K_2$  occurs on a plane ranging from  $(211)_M$  [i.e.  $(311)_A$ ] to  $(12\bar{1})_M$  [i.e.  $(3\bar{1}\bar{1})_A$ ] in the direction  $[1\bar{1}\bar{1}]_M$  [i.e.  $[0\bar{1}1]_A$ ].

For a zero value of the second shear, the predicted habit planes range from  $\{259\}_A$ , when the plane of  $K_1$  is  $(110)_A$ , to a plane close to  $\{112\}_A$ , when the plane is  $(001)_A$ . As the magnitude of the second shear is increased from zero the habit planes for all the  $K_1$  systems move towards  $\{557\}_A$  and converge on a particular habit plane with indices  $\{hhl\}$ . The exact position of the habit plane depends on the plane of  $K_2$ ; with  $K_2$  a  $(100)$  plane, and a second shear of magnitude 0.0785, the predicted habit plane is about  $2^\circ$  from  $\{557\}_A$ . The correct orientation relation is also claimed to be predicted and the magnitude

of the shape deformation ranges from 0.2 to 0.5 or higher. Kelly suggests a mechanism for accommodation of this very high shape strain.

For completeness, the transformations in Fe–Cr–Ni 18-8 type stainless steels and other steels containing appreciable amounts of chromium or manganese from f.c.c. to h.c.p. ( $\epsilon$ -phase) martensite should be mentioned. The  $\epsilon$ -martensite forms as bands on the  $\{111\}$  planes of the austenite, in very much the same manner as the transformation in cobalt already described, and there is evidence from the measured shape change that some self-accommodation by varying the shear direction also occurs. The transformation is accompanied by the formation of b.c.t. type martensite and this  $\alpha'$  phase forms as laths which are twin-related when formed away from any  $\epsilon$  phase. The  $\alpha'$  phase in isolation has a habit plane near  $(\bar{1}\bar{1}2)_A$ . Kelly has applied the BM treatment to the formation of  $\alpha'$  and finds the best agreement with  $\delta=1.018$ . The lattice-invariant strain is a shear on  $(111)(\bar{1}2\bar{1})_A$ . The variant of the habit plane predicted is nearly normal to the almost parallel planes  $(111)_A$  and  $(110)_B$ , whereas in the other solutions the habit plane has been as nearly as possible parallel to these two planes. Shimizu and Tanaka (1978) have treated the transition h.c.p.– $\alpha'$  on the assumption that the h.c.p. phase forms first, and their crystallographic results are quite similar to those of Kelly; they also had to assume a rather large dilatation to obtain the correct habit plane.

#### 94. BODY-CENTRED CUBIC TO ORTHORHOMBIC TRANSFORMATIONS

For all the transformations considered in this section, the correspondence is such that the principal axes in the product are parallel to the three edges of an orthorhombic unit cell and are derived from one cube edge and the two face diagonals perpendicular to this cube edge in the cubic unit cell. The associated principal strains are reasonably small in all cases and are consistent with the orientation relation and with the hypothesis that each lattice point of the parent structure moves directly to the nearest lattice point in the product. Moreover, in the transformations in  $\text{Cu}_3\text{Al}$  and  $\text{AuCd}$ , and in many other alloys of copper, silver or gold, both parent and product are ordered, and the postulated correspondence gives the correct ordering of the product. This is illustrated for the  $\text{AuCd}$  transformation in Fig. 22.18, where the ordered orthorhombic unit cell and one-half of the ordered b.c.c. unit cell are shown approximately in their correct orientation relationship. If three edges of the orthorhombic unit cell become equal, so that the structure becomes f.c.c., the correspondence from product to parent is equivalent to the parent–product correspondence used in the previous section.

In the  $\text{AuCd}$  transformation, the conventional unit cell of the product phase is orthorhombic, but transformations from b.c.c. to h.c.p. structures, and to other structures based on different repeat patterns in the stacking order of two-dimensional hexagonal layers, will also be discussed. It is then convenient to use an orthorhombic unit cell defined by  $\mathbf{a}_1$ ,  $\mathbf{a}_1 + 2\mathbf{a}_2$  and  $\mathbf{c}$ , where  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are vectors at  $120^\circ$  within the hexagonal layers and  $\mathbf{c}$  is normal to the layers. For h.c.p. structures, these vectors define the orthohexagonal unit

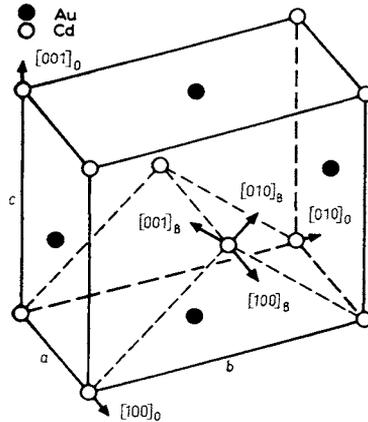


FIG. 22.18. The orthorhombic unit cell of the ordered AuCd phase (solid lines) together with one half of an ordered b.c.c. cell (broken lines) in approximately the correct orientation relation (after Mackenzie and Bowles, 1957). The b.c.c. cell is slightly distorted to simplify the diagram. The axes are labelled in accordance with the variant of the correspondence given in eqn. (93.1).

cell and the axes have lengths  $a_1$ ,  $\sqrt{3}a_1$  and  $\gamma a_1$ , where  $\gamma$  is the axial ratio  $c/a$ . This makes the second (or “ $b$ ”) axis the longest axis, which is contrary to standard crystallographic practice, but is convenient when discussing transformations involving h.c.p. structures as the  $c$  axes of conventional and orthohexagonal unit cells then coincide. For convenience in discussing the transformations together, the longest axis in the orthorhombic gold cadmium unit cell will be labelled the “ $b$ ” axis.

Denoting the parent axes by  $B$  and the product axes by  $o$  gives the following correspondence matrices

$$(oCB) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \quad (BCo) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad (94.1)$$

This is the variant used by Mackenzie and Bowles (1957) rather than that of Wechsler *et al.* (1953). As the parent structure ( $B$ ) corresponds to the product structure of the previous section, and the product structure becomes equal to the parent of the previous section if the axes are equal, it follows that  $(oCB)$  is a variant of the correspondence ( $\Lambda CM$ ) used in the last section.<sup>†</sup> The correspondence is illustrated for the case where the product is h.c.p.

<sup>†</sup>The use of different variants in this section and the previous section may be a little confusing, but is designed to facilitate comparison with the original literature. Most papers on steels, although not those of Bowles and Mackenzie, use the variant (94.1); however, the paper by Mackenzie and Bowles is the most important general treatment of the transformations in this section.

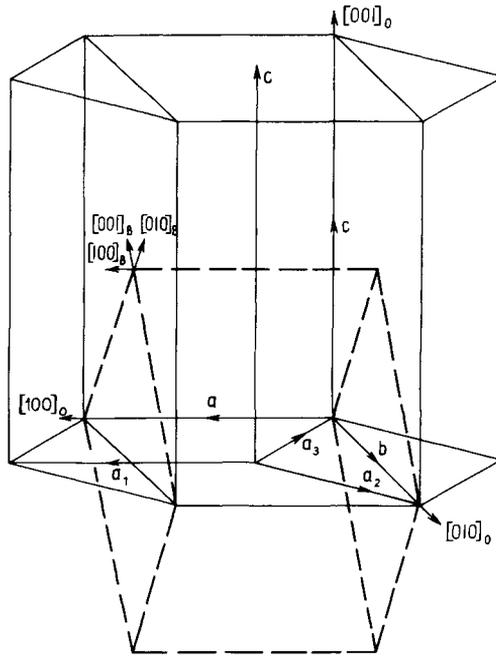


FIG. 22.19. The correspondence for b.c.c to h.c.p. transformations. The primitive h.c.p. unit cell is defined by the vectors  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{c}$  and the orthohexagonal cell by  $\mathbf{a} (= \mathbf{a}_1), \mathbf{b} (= \mathbf{a}_1 + 2\mathbf{a}_2)$  and  $\mathbf{c}$ . The broken lines show the position of the b.c.c. cell. The axes are labelled in accordance with the correspondence of eqn. (94.1).

in Fig. 22.19, which also shows the relation between the conventional (three-axis) cell and the orthohexagonal cell.

According to the assumptions made in the original Bowles and Mackenzie papers, the plane  $\mathbf{m}$  of the lattice-invariant shear is a mirror plane in the parent, and corresponds to a twinning plane  $\mathbf{m}^M$  of the product. One of the  $\{100\}$  cubic mirror planes remains a mirror plane after transformation and hence cannot be a product twinning plane. The other two  $\{100\}$  planes become  $(011)$  and  $(0\bar{1}1)$  planes of the product, and choice of either leads to equivalent predictions of the crystallographic parameters of the transformation. Two of the b.c.c.  $\{110\}$  planes also remain mirror planes after the transformation, whilst the other four become equivalent possible twinning planes of type  $\{111\}$  when referred to the orthorhombic axes. Thus with the Bowles–Mackenzie assumption, there are possibly two types of transformation utilizing different product twinning planes; these will be described as class B and class A respectively.

Consider the orthorhombic unit cell to have edges  $a, \gamma_1 a$  and  $\gamma_2 a$  and the cubic unit cell to have edge parameter  $a_0$ . The principal strains are then  $\eta_1 = a/a_0, \eta_2 = \gamma_1 a/\sqrt{2}a_0$  and  $\eta_3 = \gamma_2 a/\sqrt{2}a_0$ . When a dilatation parameter is included it is most convenient, as in the previous section, to write  $\theta = \delta a/a_0$ . The principal strains in the invariant line strain after removing the dilatation are then  $\eta_1 = \theta, \eta_2 = \gamma_1 \theta/\sqrt{2}$  and  $\eta_3 = \gamma_2 \theta/\sqrt{2}$ .

As described in Section 92, the twinning direction is fixed by the Bowles and Mackenzie assumption so that two possible classes of transformation may be defined as follows

Class A

$$\begin{aligned}(\mathbf{m}; \mathbf{B}) &= \frac{1}{2}(1\bar{1}0) \\ [\mathbf{B}; \mathbf{I}] &= [1, 1, 2(\gamma_2^2 - \gamma_1^2)/(\gamma_1^2\gamma_2^2 - \gamma_1^2 - \gamma_2^2)] \\ (\mathbf{m}^M; \mathbf{O}) &= \frac{1}{2}(1\bar{1}\bar{1}) \\ [\mathbf{O}; \mathbf{I}^M] &= [1, \frac{1}{2} - (\gamma_2^2 - \gamma_1^2)/(\gamma_1^2\gamma_2^2 - \gamma_1^2 - \gamma_2^2), \frac{1}{2} + (\gamma_2^2 - \gamma_1^2)/(\gamma_1^2\gamma_2^2 - \gamma_1^2 - \gamma_2^2)]\end{aligned}$$

Class B

$$\begin{aligned}(\mathbf{m}; \mathbf{B}) &= (001) \\ [\mathbf{B}; \mathbf{I}] &= [010] \\ (\mathbf{m}^M; \mathbf{O}) &= (0\bar{1}1) \\ [\mathbf{O}; \mathbf{I}^M] &= \frac{1}{2}[0\bar{1}\bar{1}]\end{aligned}$$

For class A transformations, the direction  $\mathbf{I}$ , or  $\mathbf{I}^M$ , is irrational. For class B transformations, the plane of shear is the mirror plane (100), and the solutions are then crystallographically degenerate (see pp. 1009 and 1016). When the product is h.c.p., the plane  $\mathbf{m}^M$  is a  $\{10\bar{1}\}$  plane (in Miller-Bravais indices) for a class A transformation and is the usual h.c.p. twinning plane  $\{10\bar{1}2\}$  for a class B transformation.

In some of this group of transformations,  $\mathbf{m}$  and  $\mathbf{I}$  have been identified by optical observations. Thus the product of the gold-cadmium reaction is a stack of orthorhombic twins, the twinning plane being observed to be that of a class A transformation, and twins of this kind have been observed in the  $\text{Cu}_3\text{Al}$  product. The results of the theoretical analysis using the above class A elements are in good agreement with the experimental results. For other metals and alloys the position is less certain, and both possibilities have to be considered.

The symmetry of the class B transformations considerably simplifies the algebra, and the crystallography may be regarded as a generalization of that appropriate to steel martensites. The analytical formulae for the habit plane, shape deformation, etc. require only slight modifications to enable them to be applicable to class B transformations (Christian, 1955-6) but they will not be repeated here. The habit plane and the shape deformation are plotted in Fig. 22.20 as a function of  $\theta^2$  for three different values of  $\gamma_2^2 = \gamma^2$ , assuming an h.c.p. product with  $\gamma_1^2 = 3$ . The variant chosen corresponds to the  $(\alpha_+ \omega_-)$  solution; the equivalent  $(\alpha_- \omega_-)$  solution is obtained by reflecting in the equator of the projection, and the  $\omega_+$  solutions are obtained by a rotation of  $180^\circ$  about the direction of projection. The habit planes do not change much with variation in  $\gamma$  and are all close to the great circles representing  $(h\bar{k}k)$  and  $(hkk)$  planes. As in the previous section, the orientation relations cannot be written conveniently in parametric form.

The class B transformations illustrate the point made on p. 1000 concerning the proportions of the two twins. It is readily seen that  $k_+^M$  and  $k_-^M$  are opposite in sign so that the macroscopic invariant plane cannot be produced by having a stack of fine  $\{001\}_0$  twins

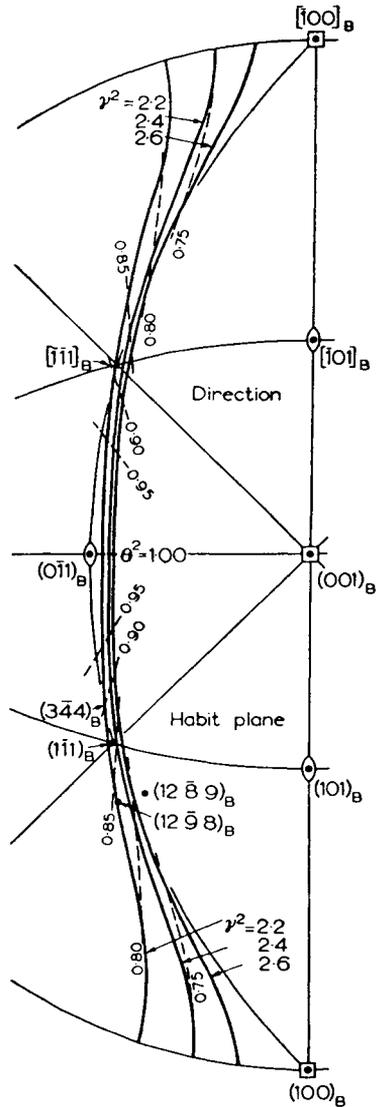


FIG. 22.20. Stereographic projection showing the variation of the predicted habit plane with  $\theta^2$  and  $\gamma^2$  for class B b.c.c. h.c.p. transformations ( $\gamma_1^2 = 3$ ) (after Mackenzie and Bowles, 1957). The numbers along the curves are 100 times the relevant values of  $\theta^2$ .

as product. This may explain why this class of transformation, which utilizes the normal twinning mode of the h.c.p. structure, is observed, if at all, much less frequently than class A transformations.

The crystallography is more complex when the class A assumption is made, and in general a numerical solution has to be obtained for each set of input data. Appropriate

computing methods were used by Bowles and Mackenzie and by Crocker and Bilby (1961). The variation of the habit plane with  $\theta^2$ , again treating the h.c.p. case with  $\gamma_1^2 = 3$ , is shown in Fig. 22.21. Once again the  $\omega_+$  and  $\omega_-$  solutions are equivalent; they correspond to reflections in the line through (001) and  $(\bar{1}\bar{1}0)$  about which the curves are symmetrical. The  $\alpha_+$  and  $\alpha_-$  solutions are not equivalent and the curve  $D - D'$  separates the diagram into two regions corresponding to these two solutions. The part of the  $\alpha_-$  solution not shown may obviously be completed by reflecting in the symmetry line.

Now compare the theoretical and experimental results, beginning with gold-cadmium. As already mentioned, this is one of only three known alloys which can be induced to transform by a single interface, and as such it has been much investigated. The relevant crystallographic parameters were measured accurately by Wechsler *et al.* (1953). The twins in the orthorhombic phase are visible under an optical microscope and are those corresponding to a class A transformation. The parameters which fix  $\eta_i$  are  $(a/a_0)^2 = 0.9007$ ,  $\gamma_1^2 = 2.3787$  and  $\gamma_2^2 = 2.2820$ .

Assuming no dilatation, the predicted habit plane for the above variant is very close to  $(3\bar{1}3)$  and the measured and predicted planes differ by less than  $1\frac{1}{2}^\circ$ . The direction of the shape deformation projected on to the habit plane is also very close to the prediction. This is illustrated by the predictions of  $\nu$  and  $\mathbf{e}$  as functions of  $\theta^2$ ; the value of  $\theta^2 = 0.90$  is well within the  $2^\circ$  circles of experimental error around the measured values. Finally, the experimental orientation relation, which is approximately  $(001)_0$  parallel to  $(011)_B$  and  $[110]_0$  parallel to  $[11\bar{1}]_B$ , is also in good agreement with the theory, according to which the angles between these pairs of planes and directions are both less than  $\frac{1}{4}^\circ$  (see p. 1024). The fraction of the twinning shear, i.e. the relative volumes of the two twins, should be 0.28 and the experimental results are consistent with this. The whole set of data constitutes an impressive agreement between theory and experiment (see Fig. 22.22).

Transformations in  $\beta$  copper-aluminium alloys were first studied in detail by Greninger (1939) and later by Swann and Warlimont (1963). On quenching from the  $\beta$  phase field, alloys with more than 11 wt.% aluminium first order to a  $DO_3$  type structure and then transform martensitically to one of two phases  $\beta'_1$  or  $\gamma'$  depending on the exact composition. Alloys with less than 11% aluminium transform to a martensitic phase  $\beta'$  which differs from  $\beta'_1$  only in not being ordered; no distinction was made between  $\beta'$  and  $\beta'_1$  before the work of Swann and Warlimont. The structure of the  $\gamma'$  phase may be described as ordered h.c.p., the true structure being orthorhombic. There has been much confusion about the structure of  $\beta'_1$ , which was described as a slight distortion of the h.c.p. structure by Isaitschev *et al.* (1938) but as ordered tetragonal by Swann and Warlimont. Later work by Wilkens and Warlimont (1963) showed that a regular arrangement of stacking faults in the  $\beta'_1$  phase gave a superstructure which is orthorhombic, in agreement with a cell found by Nakanishi (1961). There is a similar arrangement of stacking faults in the  $\beta'$  structure, which is f.c.c. if these faults are ignored.

For the  $\beta_1$ - $\gamma'$  reaction, the  $\gamma'$  plates are twinned on  $\{10\bar{1}1\}$ , so that this is a class A transformation. Using parameters given by Greninger, the predicted habit plane is approximately  $(2\bar{1}2)_B$  for the  $(\alpha_-\omega_+)$  solution and its variation with  $\theta_1^2$  for  $\gamma_2 = 1.62$  is shown in Fig. 22.23 (Mackenzie and Bowles, 1957). Two variants of the habit plane are

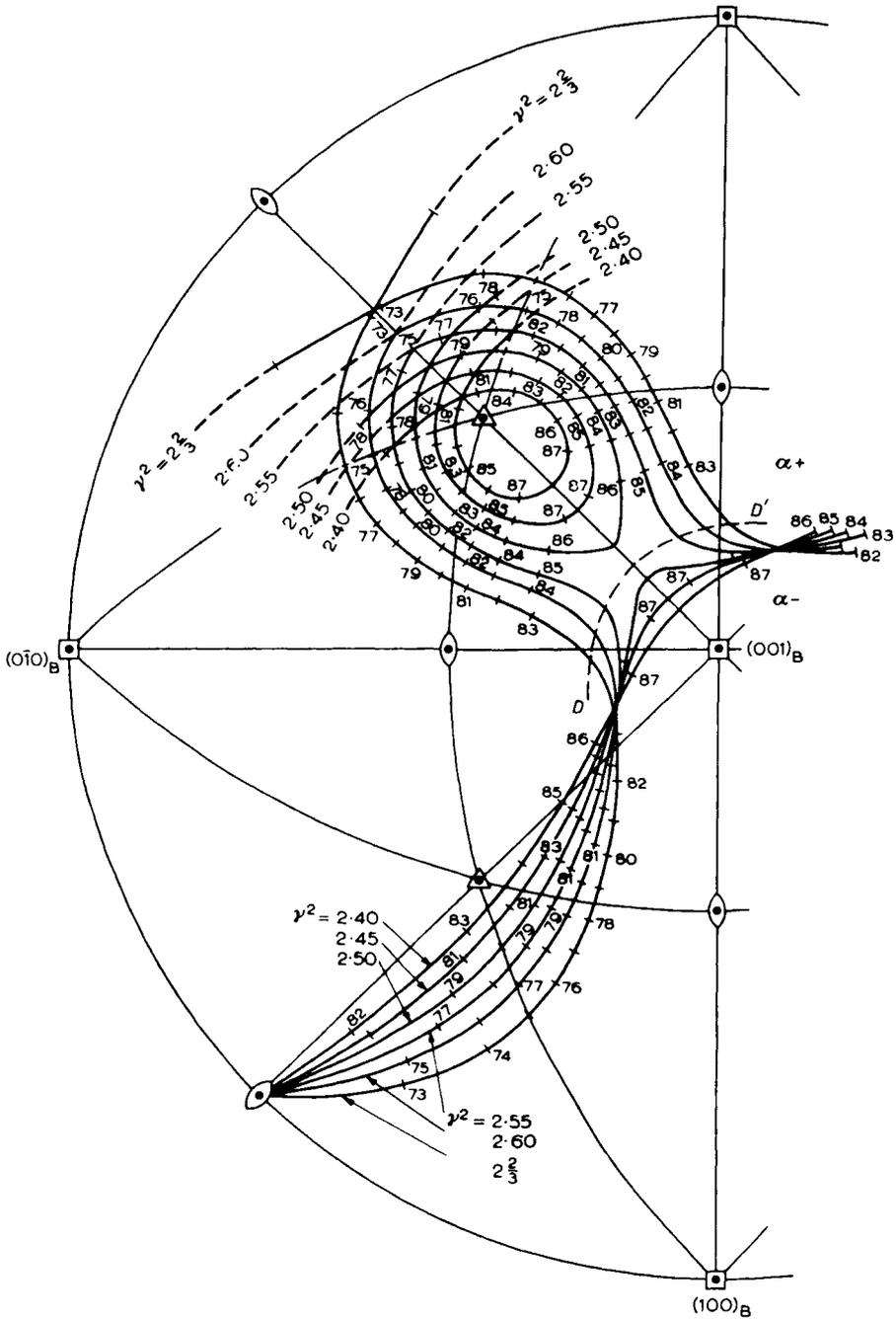


FIG. 22.21. Stereographic projection showing the variation of the habit plane with  $\theta^2$  and  $\gamma^2$  for class A b.c.c.-h.c.p. transformations ( $\gamma_1^2 = 3$ ) (after Mackenzie and Bowles, 1957). The numbers along the curves are 100 times the relevant values of  $\theta^2$ .

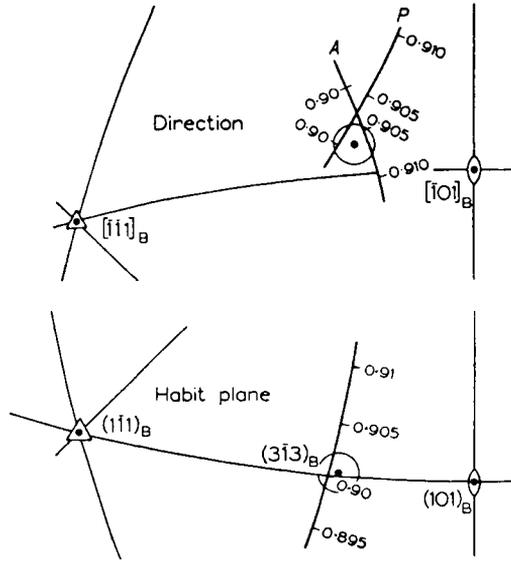


FIG. 22.22. Stereographic projection showing the predicted variation with  $\theta^2$  of the habit plane and the direction of the displacement in the shape deformation for the  $(\alpha - \omega_1)$  solution for the AuCd transformation (after Mackenzie and Bowles, 1957). The observed plane and direction (projected onto the habit plane) are surrounded by a circle of radius 2 corresponding to the experimental error. The curve labelled *A* for the direction gives the actual predicted directions, and that labelled *P* gives the projection of these directions onto the habit plane.

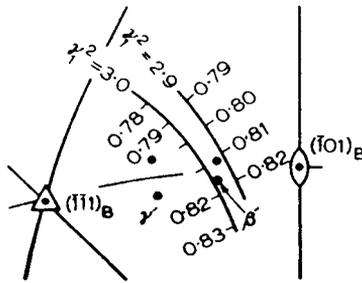


FIG. 22.23. Stereographic projection showing the predicted and measured variations with  $\theta^2$  and  $\gamma_1^2$  of the habit plane of the transformation in  $\text{Cu}_3\text{Al}$  with  $\gamma_2 = 1.62$  (after Mackenzie and Bowles, 1957). Two variants of the habit plane are plotted in the figure, together with two variants of the habit plane for the transformation to  $\beta'$   $\text{Cu}_3\text{Al}$ .

plotted, and both lie close to the  $\gamma_1^2 = 3$  curve (exact h.c.p. curve). It is not possible to distinguish between these two variants, and either possibility gives  $\Delta$  near unity as  $(a/a_0)^2 = 0.79$ . The orientation relation given by Greninger is  $(001)_0$  parallel to  $(011)_B$  and  $[110]_0$  parallel to  $[111]_B$  and this is also in reasonable agreement with the  $(\alpha - \omega_+)$

prediction. Early results for the twin ratio were about 2:1, compared with a theoretical value of 3.5:1. Using thin film electron microscopy, Swann and Warlimont obtained values ranging from 1.3:1 to 3.3:1 and suggested that some of the twins did not extend across the martensite plates.

The crystallography of the transformation to  $\beta'$  or  $\beta'_1$  was not discussed in detail by Bowles and Mackenzie because of uncertainty about its crystal structure. This transformation is of great interest because thin film electron microscopy has shown that the product structure contains arrays of stacking faults rather than twins. This is thus an example of a transformation in which the interface dislocations leave surfaces of discontinuity in the product lattice, as mentioned on p. 963. The conclusion that the faults are inherent in the transformation mechanism is supported by the observation that the fault density does not change with composition within the accuracy of the experiments, although the stacking fault energy decreases with increasing amount of aluminium. Swann and Warlimont treated the transformation as b.c.c. to f.c.c. and obtained good agreement with the experimental habit plane near  $\{133\}$  found by Greninger.

Very similar transformations occur in copper–zinc and in a large number of binary and ternary alloy systems based on copper, silver or gold. These alloys have been much studied because of their shape memory properties. All the transformations occur when the disordered phase is cooled rapidly from high temperatures. In most cases, the cooling rate is insufficient to prevent the phase ordering initially to a B2, DO<sub>3</sub> or L2<sub>1</sub> superlattice structure before transforming martensitically to a close-packed structure. The martensitic phases may all be regarded as stacking sequences of hexagonal close-packed layers of atoms. There are three basic types of martensite, namely f.c.c. (denoted  $\alpha'$ ), h.c.p. (denoted  $\gamma'$ ) and the stacking sequence...ACAB... (denoted  $\beta'$ ). To these are added the subscripts 1, 2 or 3 signifying the inherited type of order as DO<sub>3</sub>, B2 or L1 respectively. An alternative notation specifies the number of close-packed layers in a repeat unit together with an indication of the symmetry of the layer stacking as either rhombohedral or hexagonal. This leads to the product structures 3R = ...ABC... and 9R = ...ABCBCACAB... derived from B2, 6R = ...AB'CA'BC'... and 18R = ...AB'AB'AB'CA'CA'CA'BC'BC'BC'... derived from DO<sub>3</sub> or L1<sub>2</sub>, and 2H = ...AB... or ...AB'... derived from any of the ordering schemes.<sup>†</sup> The phase which actually forms on cooling varies with the composition of the alloy, the commonest being the 9R or 18R, but the 3R, 6R or 2H phases can often be induced by applying an appropriately orientated stress above the  $M_s$  temperature but below the  $T_0$  temperature at which the two phases are in equilibrium (subject to the condition of no diffusion), i.e. have equal free energies.

Few of the actual systems are truly orthorhombic because the sixfold symmetry about the  $c$  axis is destroyed by the ordering. The resultant  $c$  vector is no longer normal to the hexagonal planes but the distortion leaves a unit cell differing very little from the orthorhombic cell so that the deviation can be ignored. Treating the transformation as

<sup>†</sup>Various different notations have been used; that used by Sato *et al.* (1967) and by Warlimont and Delaey (1974) uses 1R and 3R for the structures described here as 3R and 9R.

b.c.c.–orthorhombic in all cases and ignoring the effects of atomic ordering enables the crystallographic theory, equivalent to that for steels, to be applied.

The lattice-invariant transformation is found to be  $\{001\}_0$  slip for the  $\beta'$  products and  $\{010\}_0$  twinning for the hexagonal  $\gamma'$  phases. The  $\gamma$  products thus have a class A transformation noted above whilst the  $\beta'$  products have a class B transformation. In these  $\beta'$  products, however, the lattice-invariant shear is not quite lattice-invariant inasmuch as the dislocations in the interface trail stacking faults in the product phase, thus delineating the plane of the lattice-invariant shear. The experimental habit plane for  $\gamma$  type martensites is close to  $\{122\}$ – $\{133\}$  whereas, for  $\beta$  type martensites, it is in the region  $\{133\}$ – $\{166\}$ – $\{2,11,12\}$ . Both these results are in good agreement with the predictions of the crystallographic theories as are also the measured directions and magnitudes of the shape change (see Fig. 22.23).

One interesting point arises from the heavily faulted  $\beta'$  structures. The final structure may be treated as f.c.c. and the high density of faults is then attributable to a rather large lattice-invariant deformation. Part of the lattice-invariant deformation is now producing the characteristic faults which relate the 3R structure to some higher stacking structure, and the balance of the lattice-invariant shear is visible as random stacking faults of the structure. An alternative view is that the lattice deformation is defined so that it produces the final structure whereas the lattice-invariant deformation is now much smaller and just gives the occasional stacking faults. This is a circumstance in which it is not possible unambiguously to factorize the deformation into lattice and lattice-invariant components.

The above description is all given in terms of an isolated martensite plate but, in all the copper-base alloys, it is observed that plates form in groups of four or two around a  $\langle 110 \rangle$  pole of the b.c.c. structure. These plates form a self-accommodating system so that there are almost no long-range strains associated with the transformation. In Chapter 26, it will be shown that such self-accommodating systems are very important in shape memory alloys.

The above description of a twinned product phase has simply assumed rational twin:twin interfaces, i.e. type I twinning. Otsuka (1987) has pointed out that this may be an unjustified assumption and that many transformation twins may be type II rather than type I. (This distinction does not arise in cubic metals and in certain modes of other metals where all four elements are rational, and type I and type II modes are interchangeable.) Type II twinning was first considered by Otte (1968) and by Ross and Crocker (1970). In order for the two twins to have equivalent correspondences to the matrix, it is necessary that the  $\eta_2$  axis should correspond to a twofold symmetry axis in the twin. Otsuka demonstrated that for a copper–aluminium–nickel alloy a type II twinning assumption gave much better agreement with theory than did type I twinning. A similar result was obtained for a binary copper–nickel alloy.

The remaining transformations of interest in this category are the b.c.c.–h.c.p. changes in titanium or zirconium and their alloys and the largely unexplored, similar transformations in hafnium alloys and lithium. The experimental orientation relationship differs only slightly from the simple relations suggested by Burgers in 1934. In this paper,

Burgers developed a single shear type theory for the transformation in zirconium. This theory supposes that, as the temperature of the assembly is reduced, spontaneous shears in localized regions produce the embryos.

A feature of both the titanium transformation and that of zirconium (Williams *et al.*, 1954; Gaunt and Christian, 1959) is that one of the strains is very small, which means that the lattice deformation is almost an invariant plane strain. The experimental habit plane is within  $1\frac{1}{2}^\circ$  of the theoretical prediction, and the measured shape direction also agrees well with the theoretical result for  $\mathbf{e}$ . Both titanium and zirconium are internally twinned, and are thus class A transformations. To obtain the best agreement with experiment, rather large values of the dilatation parameter had to be assumed, namely for titanium and for zirconium, but in current thinking these are probably spurious.

The experimental orientation relation is very close to the rational relation proposed by Burgers. The Burgers relation has the basal plane of the h.c.p. structure parallel to a  $\{110\}$  b.c.c. plane and a close-packed  $(11\bar{2}0)$  direction parallel to a b.c.c.  $(111)$  direction. With such a relation, there should be a  $11^\circ 32'$  angle between a second  $\langle 110 \rangle$  direction and another  $\langle 11\bar{2}0 \rangle$  direction. The Burgers orientation would thus require two variants of the h.c.p. phase, nucleated in the same b.c.c. crystal and having parallel basal planes related by an angle of about  $10.5^\circ$  about the  $c$  axis. Measurement of the actual angle thus gives a good indication of the deviation from the ideal Burgers orientation. Measurements gave values of  $9 \pm \frac{1}{2}^\circ$  for titanium and  $7\frac{1}{2} \pm \frac{1}{2}^\circ$  for zirconium.

The theory shows that there are four variants of the orientation relation with nearly parallel basal planes. For the correspondences above, the basal plane is almost parallel to  $(011)_B$  and the three  $\{11\bar{2}0\}$  hexagonal planes correspond to the planes  $(100)_0$ ,  $(\bar{1}\bar{3}0)_0$  and  $(1\bar{3}0)_0$  of the orthohexagonal cell. In the case of titanium, the four habit planes are approximately  $(\bar{1}2, \bar{8}, 9)$ ,  $(12, \bar{8}, 9)$ ,  $(\bar{1}2, \bar{9}, 8)$  and  $(12, \bar{9}, 8)$ . Labelling these as (1), (2), (3) and (4) in that order, the  $(1\bar{3}0)$  plane is within  $1^\circ$  of  $(1\bar{1}1)_B$  for variants (1) and (3) and of  $(111)_B$  for variants (2) and (4). The measured azimuthal angle between variants could then correspond, for example, to the angle between  $(1\bar{3}0)_0$  in variant (1) and  $(1\bar{3}0)_0$  in either of variants (2) or (4). The two possibilities give  $9.13^\circ$  and  $9.16^\circ$  for titanium with  $\theta^2 = 0.797$ , in excellent agreement with experiment. There are also two possibilities for the angle between the basal planes in these variants, the theoretical values being  $1.61^\circ$  and  $1.51^\circ$  and the experimental value  $1 \pm \frac{1}{2}^\circ$ .

The orientation relation is the main justification given by Mackenzie and Bowles for the conclusion that the titanium transformation is of the class A ( $\alpha_+ \omega_+$ ) type; the other possibilities all give azimuthal differences in orientation appreciably larger than that observed. Comparison of the theoretical and experimental parameters of the shape deformation is less satisfactory; the direction  $\mathbf{e}$  was not determined accurately in the work of Williams *et al.*, but the magnitude of the shape shear was found to be much less than expected. However, the discrepancy is believed to be due to accommodation slip. The direction of the homogeneous shear was measured in a titanium-molybdenum alloy by Gaunt and Christian and found to agree well with the predictions of the ( $\alpha_+ \omega_+$ ) solution. It is noteworthy that the lattice-invariant shear is only about 0.08 of the twinning shear if this

theory of titanium is correct, so that if the product is a stack of twins, one orientation should predominate. This is also in agreement with available evidence.

The rather large dilatation in the BM theory for titanium, and the even larger one for zirconium (see below), might be avoided by choice of other systems  $m$  and  $l$ . Crocker (1959) concludes from an analysis of the known twinning and slip modes that the twinning mode  $\{11\bar{2}2\}\{11\bar{2}\bar{3}\}$  is more probable than that discussed above, and also has a lower dislocation shear. However, in common with most other possibilities examined, this gives an azimuthal angle between variants which is slightly greater than  $10.5^\circ$ . In view of the evidence that there is appreciable scatter in the habit planes of individual plates (see below), it is not unlikely that several types of shear are operative, and the anomalous orientation relations might be explained in this way. However, there is no doubt that at present the BM assumption gives the best agreement between theory and experiment.

Zirconium has similar lattice spacings, and the experimental habit planes are quite close to those reported for titanium. The scatter in the results of various workers for both titanium and zirconium is illustrated in Fig. 22.24. Unfortunately, agreement between theory and experiment is not so good for zirconium. With the best available lattice parameters, a rather large dilatation of 1.5% is needed to give a habit plane near the experimental results shown in Fig. 22.24, and the predicted azimuthal difference between variants with nearly parallel basal planes is then  $10.7^\circ$  instead of the experimental  $7.5^\circ$ . The position can be improved somewhat by choosing a slightly lower axial ratio, which is not improbable because the measured  $\gamma$  might have been increased by oxygen or nitrogen contamination. Use of a large  $\theta$  also improves the agreement with the orientation relation, and reduces the magnitude of the dilatation correction, but it moves the habit plane quite rapidly away from the  $(\bar{1}\bar{1}1) - (\bar{1}00)$  boundary. It is noteworthy that the volume change for zirconium is larger than for

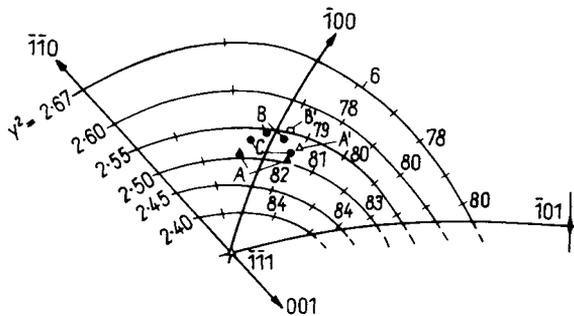


FIG. 22.24. Comparison of the predicted and experimental habit planes for titanium and zirconium. The figure shows part of the Mackenzie Bowles net showing the variation of the predicted habit plane for a class A ( $\alpha - \omega_1$ ) solution. Two variants of the experimental habit plane for titanium (Williams *et al.*, 1954) are plotted as filled triangles ( $A$ ) and the predicted habit plane with  $\gamma^2 = 2.52$  is plotted as an open triangle ( $A'$ ). Two variants of the experimental habit plane for two zirconium specimens (Gaunt and Christian, 1959) are plotted as filled circles ( $B$  and  $C$ ) and the theoretical habit plane with  $\gamma^2 = 2.55$  is plotted as an open square ( $B'$ ).

titanium, so that accommodation stresses may play a larger role in modifying the ideal crystallography and producing variations from plate to plate. If such variations exist, the discrepancies are not surprising.

Cubic-hexagonal transformations of a martensitic nature also occur in many titanium and zirconium alloys when the primary solid solutions in the b.c.c.  $\beta$  phases are either rapidly cooled or deformed. In titanium-manganese and titanium-molybdenum alloys, two different kinds of transformation have been observed, and are usually characterized by habit plane indices of  $\{334\}$  and  $\{344\}$  respectively (Liu and Margolin, 1953; Weing and Machlin, 1954; Liu, 1956; Gaunt and Christian, 1959). Actually, habits of both kinds may scatter right across the corner of the unit triangle near the  $\{111\}$  pole, and the real distinction only appears when the shear direction of the shape deformation is also determined. Figure 22.25 shows a group of  $\{344\}$  habits in a titanium-12.5% molybdenum alloy, together with the corresponding shear directions (i.e. the projections of the direction  $e$  on the habit plane). The shear directions all lie close to the great circle through  $(111)$  and  $(011)$ . Figure 22.26 shows a group of habits in an alloy containing 11% molybdenum, all but one of which are of  $\{334\}$  type with shear directions lying near the great circle through  $(111)$  and  $(001)$ . The conditions under which the two types of habit occur are not clear, but it is established that they may form in the same specimen.

The  $\{334\}$  habit is similar to the transformations in pure titanium and zirconium, and need not be further discussed; as is shown in Fig. 22.26, the experimental shear direction agrees with the theory. For the  $\{334\}$  habits, there is conflicting evidence on the orientation relations, and it is not possible to give a satisfactory theoretical discussion at present. Most of the experimental results are consistent with the assumption that the transformation is either a class A ( $\alpha\omega$ ) transformation, or a class B transformation. In both cases, however, a rather large dilatation is required and there are other difficulties.

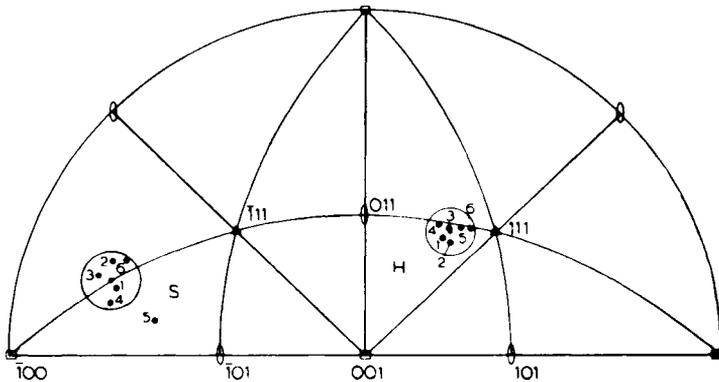


FIG. 22.25. Habit plane ( $H$ ) and direction of the shape shear ( $S$ ) for stress-induced martensite in a titanium-12.5% molybdenum alloy (after Gaunt and Christian, 1959). The shear magnitudes were: (1) 0.268; (2) 0.272; (3) 0.283; (4) 0.22; (5) 0.245; and (6) 0.20. The small circles at  $H$  and  $S$  have radii of  $6^\circ$  and  $9^\circ$  respectively.

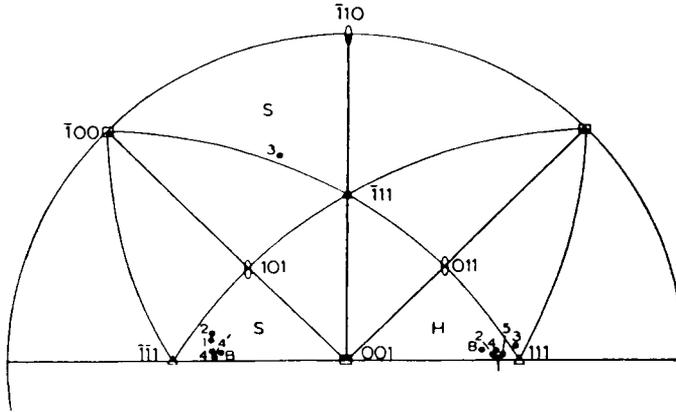


FIG. 22.26. Habit plane ( $H$ ) and direction of the shape shear ( $S$ ) for stress-induced martensite in a titanium 11% molybdenum alloy (after Gaunt and Christian, 1959). System 3, which is of  $\{344\}$  type, has been plotted in the adjacent unit triangle for greater clarity. 4 and 4' represent alternative directions derived from different tilts measured on one set of plates and  $B$  gives the theoretical habit plane and tilts derived for pure titanium according to the Bowles Mackenzie theory.

## 95. OTHER APPLICATIONS OF THE FORMAL THEORY

This section describes some other applications of the formal theory of crystallography in outline only in view of the detailed comparisons of theory with experiment made in the last two sections. A description is first given of the transition from a high temperature f.c.c. structure to a low temperature f.c.t. structure which is observed in indium-rich alloys and in quenched metastable manganese-rich alloys. The axial ratio of the product is greater than unity in the indium alloys and less than unity in some of the manganese alloys, but the crystallography of the transitions is equivalent in the two cases,  $|(c/a) - 1|$  being very small.

When referred to the conventional cubic and tetragonal cells, the lattice correspondence matrix of this transformation is the unit matrix  $I$ . The principal strains are then all very small and are approximately  $2\varepsilon$ ,  $-\varepsilon$  and  $-\varepsilon$ , assuming no volume change, where  $3\varepsilon = (c/a) - 1$  and terms of order  $\varepsilon^2$  are assumed to be negligible. This transformation may be regarded as a special case of the steel transformation discussed in Section 92 with  $\gamma$  very close to  $\sqrt{2}$ , and the treatment given there shows that the habit plane is very close to a  $\{110\}$  plane of the f.c.c. lattice. The essential features of the transformation were described by Bowles *et al.* (1950) before the final development of the formal theory of martensite crystallography. They showed that if the cubic structure is given (mathematically) two successive shears on  $\{110\}$  planes at  $60^\circ$  to each other, the tetragonal structure is obtained. (Strictly, this gives a triclinic lattice, but it is tetragonal if terms of order  $\varepsilon^2$  are ignored.) It follows that  $\{110\}$  planes of the parent phase are close approximations to invariant planes

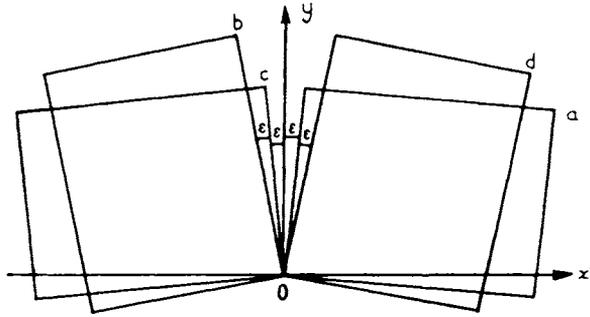


FIG. 22.27. Lattice orientations in indium–thallium produced by a “first” shear of  $(101)[101] 2\varepsilon$  (after Basinski and Christian, 1954). The “second” shears are: (a)  $(110)[\bar{1}\bar{1}0] 2\varepsilon$ ; (b)  $(110)[\bar{1}\bar{1}0] 4\varepsilon$ ; (c)  $(1\bar{1}0)[110] 2\varepsilon$ ; and (d)  $(1\bar{1}0)[\bar{1}\bar{1}0] 4\varepsilon$ .

of the transformation, and the observed habit planes are indeed indistinguishable from  $\{110\}$ .

Although this theory is not exact, it is instructive to examine its prediction in a little more detail. A “first” shear of magnitude  $2\varepsilon$  in the  $[10\bar{1}]$  direction on the plane  $(101)$  can give four different tetragonal orientations as shown in Fig. 22.27. These are produced by “second” shears of  $2\varepsilon$  and  $4\varepsilon$  respectively on  $(110)$ ,  $\pm[\bar{1}\bar{1}0]$  and  $(1\bar{1}0)$ ,  $\pm[110]$ , the oppositely directed shears on a given plane leading to twin-related orientations. These results are specified on the assumption that  $c/a > 1$ ; for  $c/a < 1$ , the second shears would be on  $(011)$  and  $(01\bar{1})$ . Thus one martensite plate with a shape deformation represented by the  $(101)[10\bar{1}]2\varepsilon$  shear can contain four product orientations in two twin-related pairs. A plate with the opposite shape deformation  $(101)[\bar{1}01]$  can also contain four orientations, these now consisting of two twin pairs on  $(011)$  and  $(01\bar{1})$ . Thus for each  $\{110\}$  habit plane there are eight orientations of the tetragonal product. These orientations are all different (but only slightly so) from the orientations produced in the plates on the  $\{110\}$  plane at  $90^\circ$  but there is one common orientation in the sets produced by planes at  $60^\circ$  to each other. Each cubic crystal may thus give rise to 36 tetragonal orientations. Experimental pole figure determinations verify these approximate predictions, and two sets of twins with the predicted crystallography are visible in each “main band” under the optical microscope.

The exact theory of the transformation for a particular indium–thallium alloy predicts a habit plane of type  $\{1, 0.013, 0.993\}$  which differs from the rational  $\{101\}$  plane by about  $20'$ . There are 24 variants of this habit plane, each of which may contain two twin-related orientations of the product lattice so that there are 48 orientations. The approximate solution is thus a degenerate case of the general theory having a smaller multiplicity of solutions. The predicted habit planes are grouped in fours around each  $\{110\}$  plane, two of them giving a shape deformation of one sign and two of the other. Similarly, the difference between the 36 orientations of the approximate solution and the 48 of the general solution arises because main bands at  $60^\circ$  to each other contain two orientations which differ infinitesimally from each other.

Observations of this transformation, both in polycrystalline specimens and in single crystals, have verified the crystallographic theory with considerable precision. The plates or main bands are parallel-sided rather than lenticular, even in polycrystalline specimens, and this almost certainly results from the degeneracy of the shape deformation as adjacent bands with the same  $\{110\}$  habit plane may have opposite shape deformations. The transformed regions thus form a self-accommodating system, and complete transformation is possible within a small temperature range. The self-accommodation and the small magnitude of the shape deformation ensure that there is no plastic deformation of the matrix and the transformation is fully reversible. In both indium–thallium and copper–manganese alloys (Basinski and Christian, 1951–2) the same transformation microstructure may be obtained repeatedly by cycling through the transformation range.

An interesting observation, first made by Bowles *et al.* (1950), is that the main bands at  $60^\circ$  to each other form interpenetrating patterns, the penetrating bands gradually tapering, whilst those at  $90^\circ$  do not interpenetrate. This observation is clearly related to the existence of a common orientation in the two sets of  $60^\circ$  bands, but it does not imply growth from a common nucleus. Observations show that genuine interpenetration of separate sets of bands occurs, this being the result of compatibility conditions of the shape deformation (Basinski and Christian, 1956); one set of bands is able to accommodate the shear strain produced by the other set.

The crystallographic theory shows that the lattice-invariant deformation may be accomplished by fine twinning of the product in the volume ratio of 2:1, and that this is converted into a ratio of 1:2 when the product is subjected to a  $60^\circ$  macroscopic shear of the other interface. These interfaces, termed  $X$  and  $X'$  by Basinski and Christian, produce respectively a single crystal product (in the orientation common to the two sets) and a twinned product with a reversed volume ratio (see Fig. 21.18).

The varied modes of transformation in single crystals of the parent phase provide excellent illustrations of the principles of compatibility of the macroscopic shape deformations. Some of the interfaces in Fig. 21.18 are genuine interphase interfaces, whilst others are just boundaries between different sets of variants of the product phase. It is also possible to have interfaces which are diffuse; the interfaces being produced by gradual tapering of a set of twins until eventually they all disappear. The interfaces can be considered as surface dislocations and the crossing interfaces then meet at surface dislocation nodes; a more detailed discussion is given by Basinski and Christian (1954, 1956).

The very small divergence between the predictions of the exact crystallographic theory and those of the approximate theory is not experimentally detectable in indium–thallium and manganese–copper alloys, but it is of considerable interest that a similar but slightly larger deviation can actually be studied in the cubic to orthorhombic order–disorder reaction in the equiatomic gold–copper alloy. Although this is an ordering reaction and must involve atomic migration, a tilt of the surface is clearly observed as the ordered regions grow at the expense of the cubic matrix. This is thus one of the reactions where a correspondence is maintained during thermally activated growth (see pp. 961–962) and to

which the theory of martensite crystallography applies. A detailed theoretical and experimental study by Bowles and Smith (1960) shows that ordered plates grow in groups of four, appearing on the surface as roughly diamond-shaped figures with four non-parallel variants of the same habit plane which is near  $\{110\}$ . The correspondence of the transformation is the same as that for the cubic-tetragonal change and the plates are internally twinned on  $\{101\}$  planes of the orthorhombic structure. The predicted results were in excellent agreement with the habit plane, orientations and shape deformation measured experimentally but it was necessary to assume a value of the dilatation parameter,  $\delta$ , of 0.9986 in order to obtain this good agreement. Whether or not this means that the concept of a dilatation is a true one is difficult to decide.

From a general point of view, it is clear that the growth of an ordered structure requires atomic migration but no long-range diffusion. As discussed earlier, the important factor here must be the maintenance of the correspondence and the associated shape change rather than the glissile nature of the interface. It is noteworthy that the change from the cubic phase to the ordered tetragonal phase CuAuI which takes place below 280 C in the same alloy is apparently not related to martensitic transformation and does not proceed by the formation of tetragonal crystals within the cubic matrix. Banded structures are found in this case also, but apparently result from mechanical twinning of the already ordered matrix. This is rather remarkable when it is remembered that the orthorhombic CuAuII structure which forms at slightly higher temperatures is a kind of superlattice or regular antiphase domain structure of the tetragonal phase. Similar reactions occur in cobalt platinum, magnesium cadmium and other alloys (Bowles and Malin, 1960).

The transformation from  $\beta$  to  $\alpha$  uranium in dilute uranium alloys is believed to be martensitic in nature but, in view of the complex crystal structure, only one atom in 30 (at best) can be carried to its correct position by the lattice deformation; the others must also shuffle. In view of the complex shuffles, it is not surprising that the growth process appears to be thermally activated. The transformation differs from all others described in several ways. It is the only martensitic transformation known in which the principal axes of the lattice deformation are not parallel to simple low index directions of the parent and also the only transformation in which there is not a unique correspondence.

An experimental study of the crystallography of the transformation in a 1.4 at.% Cr alloy (Butcher and Rowe, 1956) showed the existence of two orientation relations, in both of which  $(140)_b // (021)_a$  and  $(001)_b // (124)_a$ . In the first relation  $(001)_a$  is parallel to  $\{817\}_b$  and in the second relation it is parallel to  $\{212\}_b$ . These two related orientations both have a habit plane near  $\{321\}_b$ . Pole figures also suggested the existence of a third orientation in which  $(001)_a // (410)_b$  and this is believed to be a  $\{112\}_a$  twin of the first relation. In a systematic search for a lattice correspondence in which a primitive  $\beta$  cell would constitute a unit of transformation, Lomer (1956) found only one possibility out of some 1600 examined in which the length changes were less than about 10% and this seems to account for the first orientation above. Two other correspondences, both involving two  $\beta$  cells as the unit of transformation, appear to be related to the other two observed orientation relations, but these were discovered by using the experimental results as a guide.

The unusual correspondences, together with the lack of information on a suitable lattice-invariant shear system, make the full calculation of the crystallography of this transformation a formidable task. There are many known modes of deformation of both  $\alpha$  and  $\beta$  uranium and inclusion of even the most likely of these would require much computing time. It is thus not surprising that no detailed analysis of this transformation has been published, although some systematic calculations have been made by Crocker (1959).

#### 96. THE MARTENSITE INTERFACE AND THE FINE STRUCTURE OF MARTENSITE

The displacement of an interface between two crystalline regions which differ in orientation or structure is conservative if all the atoms in the first crystal swept by the boundary are incorporated into the second crystal; it is non-conservative if some atoms have to be added or removed from the swept region in order to facilitate this event. Conservative interfaces may be termed glissile, in the sense in which this term is used of dislocation lines. The macroscopic properties of such an interface are entirely specified by the direction of the normal  $\mathbf{v}$  to its surface and the shape tensor  $\mathbf{E}$  which defines the deformation as the interface migrates. Suppose the interface is planar and migrates in the direction  $\mathbf{v}$  without any accompanying movement of lattice dislocations. Provided the two crystals remain in contact at the migrating interface, the shape change  $\mathbf{E}$  must represent an invariant plane strain or, more accurately, it can differ from such a condition only by small elastic strains.

Following the early models of tilt and twist boundaries, Frank (1953) first showed how any grain boundary may be ascribed a dislocation content in which the net Burgers vector of the dislocation lines crossing any vector in the interface could be specified. Such dislocations have been called "primary", "intrinsic", "mathematical" and "anticoherency" and the Burgers vectors are defined in any reference lattice from which the two real lattices are defined by appropriate affine transformations. Frank's theory was generalized by Bilby (1955, 1956), who showed that a net dislocation content may be defined for a general interphase interface, and Bollmann (1970) obtained a relation in his theory of the O-lattice (see pp. 366-370) which is essentially equivalent to the main equation of the Frank-Bilby theory (Christian, 1975).

Bilby introduced the concept of the surface dislocation characterized by a surface dislocation density tensor. This tensor provides a convenient (although unfortunately not unique) description of a defect which, like a dislocation line, is subject to topological constraints and cannot end within a crystal. However, three or more surface dislocations may meet along an interior line of the crystal and the compatibility condition on this surface dislocation node may be compared with the conservation of Burgers vector at a dislocation node.

The Frank-Bilby equation arises from the incompatibility of the shape and lattice deformations. In Bilby's original derivation a three-dimensional continuous distribution of dislocations, without accumulating long-range stresses, specified by the dislocation

density tensor  $\alpha^{ij}$  introduced by Nye (1953), is confined to a sheet of thickness  $t$ , and then  $t$  is allowed to tend to zero whilst  $\alpha^{ij}$  tends to infinity in such a way that the product  $t\alpha^{ij}$  remains finite. This gives a surface dislocation separating two dislocation-free lattices, but the result can be obtained much more simply by using a circuital definition as in Figs. 8.9 and 8.17. If the two real lattices are generated by deformations  $S_A$  and  $S_B$  from a reference lattice, then the total Burgers vector, defined as the closure failure in the reference lattice of a closed circuit in the real bicrystal, crossing a vector  $\mathbf{p}$  in the interface is given by

$$\mathbf{b}_T = (\mathbf{S}_A^{-1} - \mathbf{S}_B^{-1})\mathbf{p} \quad (96.1)$$

where  $S_A$ ,  $S_B$  are the lattice deformations generating A and B from the reference lattice. This is the Frank–Bilby equation, but it is often useful to use one of the crystal lattices as the reference lattice, so that the equation becomes

$$\mathbf{b}_T = (\mathbf{I} - \mathbf{S}^{-1})\mathbf{p} \quad (96.2)$$

referred to the A lattice, or equivalently

$$\mathbf{b}_T = (\mathbf{S}^{-1} - \mathbf{I})\mathbf{p} \quad (96.3)$$

referred to the B lattice, where  $S$  is the deformation carrying A into B. The equation in any of the three forms gives the effective dislocation content of the interface, but it is not unique because of the multiplicity of affine deformations which will generate one lattice from another. If the interface is an invariant plane of  $S$ , the dislocation content is zero. It might be considered that the description which gives minimum dislocation content is the most physically significant, but consideration of a symmetrical tilt boundary shows that this is not universally true. A symmetrical tilt boundary may be produced by a simple shear on the interface plane, in which case the dislocation density is zero, but for small angles of tilt, it is usually considered physically more correct to regard  $S$  as a pure rotation produced by an array of edge dislocations. Note that, in such a case, it is possible to call the boundary fully coherent (though on a high index plane) or semi-coherent, and the sharp distinction between the two kinds of boundary no longer exists.

It should be emphasized here that Bilby's theory is essentially an appropriate continuum theory in which individual dislocations play no role. It specifies the total Burgers vector density in any direction of the interface.

Olson and Cohen (1976) and Bonnet (1981abc, 1982, 1983) have developed an approach to the elastic fields of interfaces in terms of dislocation distributions. The basis of this description is to use two arrays of dislocations of opposite nature and with Burgers vectors of opposite sign. When the interface has no long-range stress field, the short-range elastic field is modelled by two cancelling arrays (see Fig. 22.28). One array may be regarded as stress-generating; it has a uniform distribution of Burgers vector and there are no discontinuities across the interface, which is apparently fully coherent. For this reason, Olson and Cohen called this an array of "coherency" dislocations.

To eliminate the long-range stress field of such an interface which is elastically strained into full coherence, a second set of dislocations is introduced. This set consists of "real" dislocations with lattice Burgers vectors which may be defined by Burgers circuits, and the

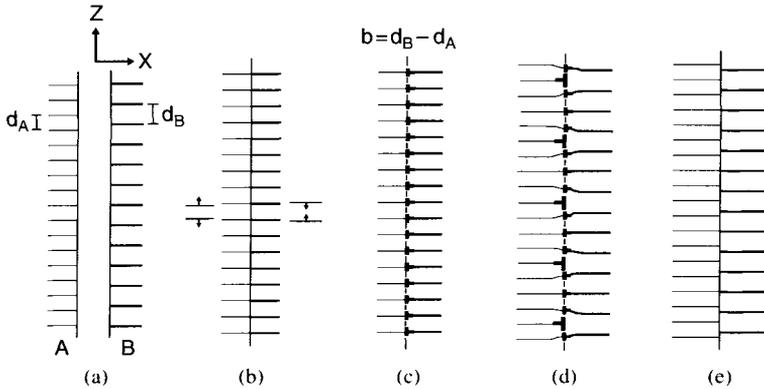


FIG. 22.28. Olson and Cohen's definition of coherency and antioherency dislocations. Interphase boundary with uniaxial distortion: (a) two crystals, *A* and *B*, with different interplanar spacings,  $d_A$  and  $d_B$ , in the  $z$  direction; (b) position of planes across coherent interface; (c) partial dislocations required in discrete lattice to maintain configuration of (b); (d) semi-coherent boundary with antioherency dislocations superimposed on coherency dislocations of (c); (e) incoherent boundary.

net Burgers vectors of which exactly cancel the Burgers vectors of the continuous virtual dislocations of the first set. The Frank–Bilby equation gives the Burgers vectors of either set (with a change of sign) and thus may be used to specify the spacing of antioherency dislocations needed to convert a given plane into an undistorted plane.

Another way of looking at this problem of the dislocation structure of an interface is to consider the definition of zero dislocation density along the interface. This may be taken either as a condition in which the interface has no long-range stress field, in which case the opposing effects of coherency and antioherency dislocations may be included, or it may with equal validity be taken as one where there are no lattice discontinuities as revealed by Burger circuits, in which case only the antioherency dislocations are real. The coherency dislocations are really simply a device to model the elastic field by a dislocation distribution, as is often done in fracture theory where a crack is modelled by a dislocation array.

The formal theories of martensite crystallography lead naturally to the concept of the martensite interface as a glissile array of dislocations, motion of which will produce the lattice-invariant deformation. If the interface contains a single set of parallel dislocations, it will be glissile if these dislocations have a Burgers vector which is a lattice vector (defined in either lattice) and which is not in the plane of the interface, except for pure screws. The first specific model of such an interface (Frank, 1953) appeared before the formal theories, and was applied to the case of  $\{225\}$  martensite in steels. As this habit plane contains the close-packed  $\langle 110 \rangle$  direction of the product lattice, Frank suggested that close-packed planes of the two lattices meet edge to edge in the interface along close-packed rows. A suitable macroscopic shear of about  $(32)^{-1/2}$  will correctly adjust the spacing between the close-packed rows in the habit plane but the whole shape deformation

cannot be an invariant plane strain because the lattice parameters are such that the interatomic distances along the close-packed rows are not the same in both crystals. These distances have to be equalized by allowing some distortion in the habit plane interface and Frank used a different assumption from the uniform dilatation later adopted by Bowles and Mackenzie. Finally, to make the atomic arrangements in the close-packed rows match properly a screw dislocation with a lattice Burgers vector has to be introduced every sixth atomic close-packed plane. The structure of the interface is shown diagrammatically in Fig. 22.29. These dislocations are “anticoherecy” dislocations in Olson and Cohen’s nomenclature because they convert a fully coherent interface with a long-range stress field into a semi-coherent interface with no long-range stress. The motion of these screw dislocations produces the lattice-invariant deformation of the formal theories.

The equivalence of the lattice-invariant deformation and the twinning assumption of the formal theories means that, for each model of the interface of the type shown in Fig. 22.29, in which the slip plane and direction correspond to possible twinning elements of the interface, there is a corresponding model in which the interface contains no dislocations, but in which the product is twinned. Figure 22.30 shows the relation between the two models so far as the structure of the interface is concerned. The displacement produced by the Burgers vector of the dislocations is now spread over the widths of the twins; the average shear depends on the Burgers vector and spacing of the dislocations, or on the twinning shear and volume fraction of the twins.

There is now overwhelming evidence that, in many martensitic transformations, the product is actually a stack of fine twins, so that the dislocation model of the interface does not apply. Twins were first detected in transformations like that occurring in indium-thallium alloys where, as already noted, they are visible with an optical microscope. These very wide twins do not exist right to the interface, however, as the atomic mismatch would

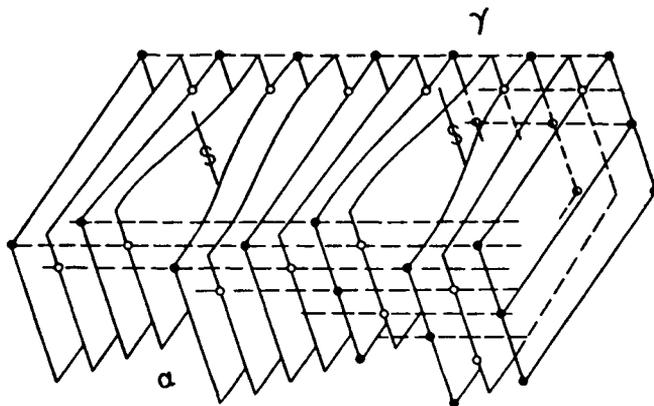


FIG. 22.29. Model of the austenite-martensite interface with close-packed planes meeting edge to edge along a close-packed direction within the interface (after Frank, 1953). The vertical sheets are  $(111)_A$  and  $(101)_M$  planes; a slight bend at the surface of contact is omitted from the figure.

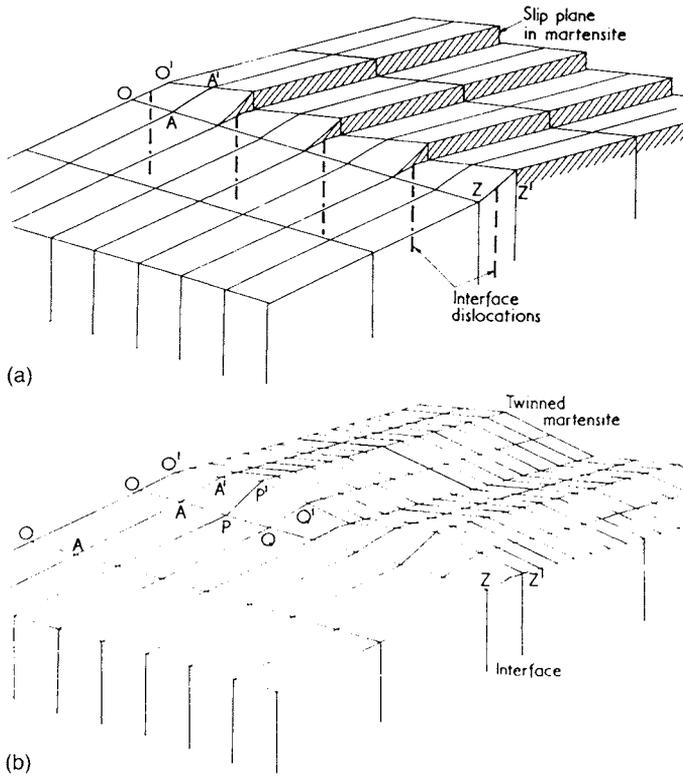


FIG. 22.30. Diagram illustrating the macroscopic equivalence of the dislocation and twin models of the martensitic habit plane interface. The slip or twinning direction is assumed to lie in the interface. In (a) the matching of the two structures is effected by a set of screw dislocations, as in Frank's model. In (b) the matching is accomplished by having a product which consists of a set of fine parallel twins.

then become large and the corresponding strain energy appreciable. There are two opposing factors which presumably govern the width of the twins. The smaller the width, the smaller will be the strain energy caused by the misfit within individual twin regions, and this will give an energy term approximately proportional to the square of the width. The second factor is the energy of the coherent twin boundaries in the product crystal, and this will be proportional to the reciprocal of the width. If the energy is minimized, there will be some width which gives the most favourable interface structure. It is improbable that the twins will be sufficiently wide to give atomic misfit of more than about one interatomic distance.

A further reduction in energy may be possible if the twins are very narrow in the interface region, but widen to many thousands of interatomic distances behind the interface. The strain energy associated with the wide twins is then reduced by being distributed over an accommodation volume behind the interface, whilst the high density of

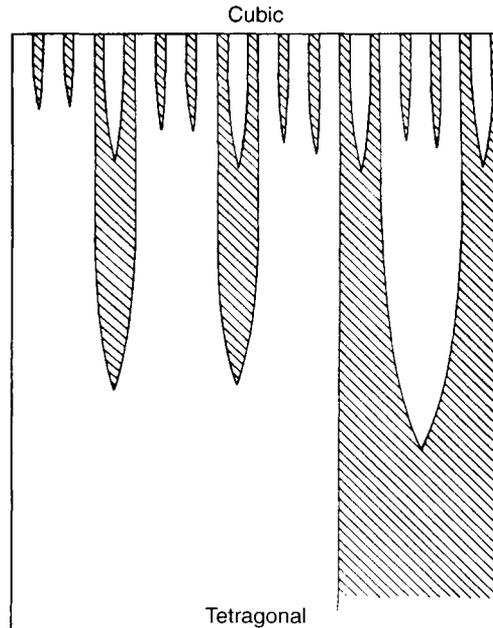


FIG. 22.31. Suggested mechanism for thickening of sub-microscopic twins in an accommodation region behind the interface (after Basinski and Christian, 1954).

twin interface is present only in a relatively small volume of the product. A possible mechanism for increase in twin width behind the interface is shown in Fig. 22.31; on an atomic scale, the tapering of the twins is accomplished by individual steps (twinning dislocations) which must therefore move along behind the true interface. Some such mechanism apparently operates in indium-thallium alloys as photomicrographs of single-interface transformations show that the twins are readily visible some distance behind the interface but fade out in the immediate vicinity of the interface. When the boundary is moved very rapidly into the parent region, the accommodation gap widens very appreciably, and then narrows again, indicating that the arrays of twinning dislocations are not as mobile as the interface itself.

Fine parallel twinning of the type observed in indium-thallium alloys is not visible when martensite formed in steels or in most non-ferrous alloys is examined under the optical microscope. This is merely a matter of resolution, and many investigations by electron microscopy have shown that the structure of martensite is indeed frequently duplex and consists of stacks of very fine twins. Two kinds of electron microscope investigation have to be distinguished. In the first of these (Pitsch, 1958-9), the martensite is formed by the heat treatment of previously prepared thin foils, whereas in the second (Kelly and Nutting, 1960-1, 1961), the martensite is obtained by heat-treating bulk specimens which are subsequently used to prepare thin foils for examination in the microscope. More recently,

the techniques of high resolution electron microscopy, which require very thin specimens, have been used to produce lattice imaging pictures of the structure of the interface itself. Directly transformed thin foils produce martensite which is not subject to the same constraints as martensite formed in bulk specimens and results will be mainly quoted from the bulk specimen type of experiment. The early work of Kelly and Nutting, Shimizu (1962), Nishiyama and Shimizu (1962) and Warlimont (1962) first showed the twinned structure of the martensite plates and also indicated that narrow striations within the plates represented bands of twins (see Fig. 22.32). This early work also verified the crystallography assumed in the standard theories of Section 92 and, in particular, the twins were found to be on the correct variant of the twinning plane relative to the habit plane and orientation relation.

Later work has shown many variations in the substructure of martensite in steels, but observed twins are always found to be on the particular twinning planes predicted by the crystallographic theories. The most convincing electron micrographs are those corresponding to the martensite structures which give best agreement with the theory, i.e. so-called thin-plate martensite. Figures 22.13 and 22.33 show the twinned structure of  $\{3,10,15\}$  martensite plates in which the twin distribution is uniform, the macroscopic interface is very planar and the twins extend through the whole cross-section of the plate.

A rational, fully coherent interface migrates by the motion of steps, i.e. transformation dislocations along the habit plane interface. In the case of the h.c.p. f.c.c.



FIG. 22.32. Early transmission electron micrograph of a 1.0% carbon steel quenched to produce martensite (courtesy of Kelly and Nutting, 1960 1). Twins having a minimum thickness of 2 nm are visible in the martensite.

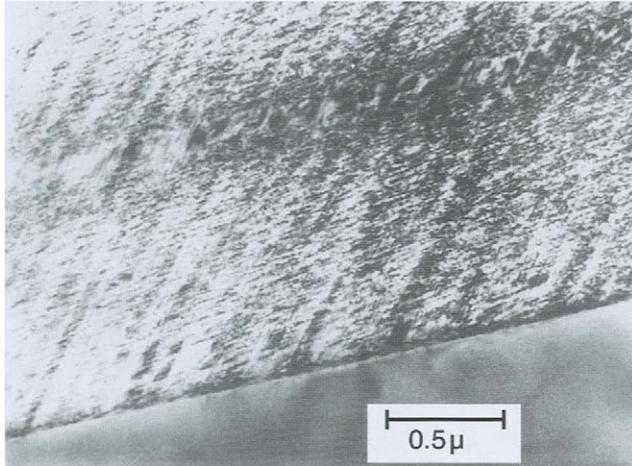


FIG. 22.33. Finely twinned structure of  $\{3, 10, 15\}$  thin-plate martensite in iron-nickel-carbon alloy to illustrate planar macroscopic interface.

transformation on heating in cobalt-nickel alloys, an experimental study by Hitzenberger and Karnthaler (1991) shows that the cubic phase grows from the thick into the thin regions of the foil as lamellae with a thickness of six to ten atom layers. There is little or no long-range stress field associated with the lamellae and it is thus believed that each lamella contains the equivalent of three, or a multiple of three, unit height steps in which are contained all three possible dislocations with shear components equal to those of Shockley partials. The minimum step height thus corresponds to six atomic plane spacings of the  $\{111\}$  cubic planes (the  $\{0001\}$  h.c.p. planes). Similar results were obtained by Waitz and Karnthaler (1997) for transformations in both directions of cobalt-iron alloys where the structure change is from f.c.c. to d.h.c.p. As described in Section 89, somewhat different results were obtained from the cooling transformation in cobalt-nickel alloys where both Gaunt and Christian (1959) and Waitz and Karnthaler (1997) found that elementary steps of the same Burgers vector produce a macroscopic shear which only alternates on a scale of several hundred atom planes. Figure 21.5 shows the free surface of a cobalt-nickel single crystal transformed from cubic to hexagonal, in which the three different surface tilts can be readily distinguished; Fig. 21.6 shows a more recent example in which the surface tilts were distinguished by atomic force microscopy.

As mentioned above, in copper-aluminium and very probably other copper alloys, the substructure consists of an array of parallel stacking faults (see Fig. 22.34) and this indicates that the lattice-invariant shear has been produced by partial dislocations in the interface. If these partials are combined in pairs to make lattice dislocations there will be no discontinuities in the martensite structure and this is the usual model of single crystal martensite. However, if several such partial dislocations are located on adjacent

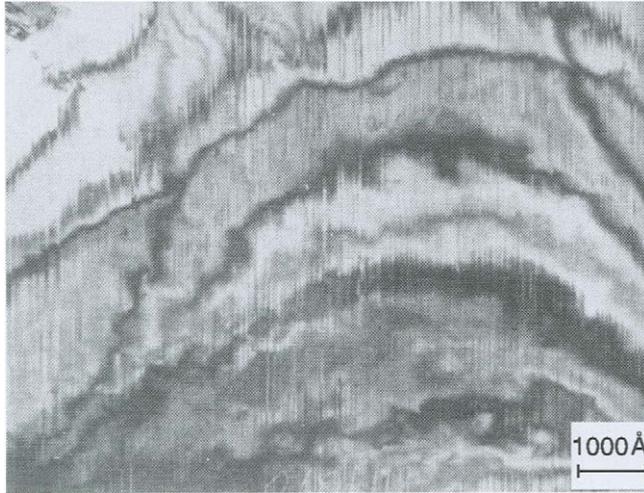


FIG. 22.34. Stacking faults in martensite formed in a copper 12.2% aluminium alloy quenched from 1199 C (courtesy of Swann and Warlimont, 1963).

planes, they will produce a twin of the lattice in which they were defined. Thus in modelling martensite there may be a hierarchy of structures ranging from isolated stacking faults to twins.

Electron microscopy, however, although verifying many of the basic assumptions of the theory, including the specific variant of the twinning elements, also raised many new problems. Reference has already been made to the uneven distribution of twins within the martensite plate and to the fact that lath or needle products tend in general to be untwinned but to contain high densities of dislocations. Figure 22.14 shows twins confined to a narrow mid-rib region of an iron–nickel alloy plate with a very irregular outer surface. The macroscopic shape deformation in such cases is apparently the same in the twinned and untwinned regions. It has been suggested that the twinned portion formed first at the relatively low temperature of nucleation but then recalescence raised the interface temperature into a region where growth by dislocations was preferred. Figure 22.35 shows the structure of a  $\{225\}$  type plate in which a band of twins has formed along one side of the plate only. Structures of this type provide the main justification for inclusion of at least a component of the twinning shear in a multi-shear description of  $\{225\}$  martensite and Fig. 22.17 shows the microstructure of a low carbon steel with a high dislocation density inside the individual laths or needles.

The structure of interfaces at an atomic level may be studied by computer simulation or by high resolution electron microscopy. Computer modelling is a much more reliable method of investigating atomic structures, now that realistic interatomic interactions are

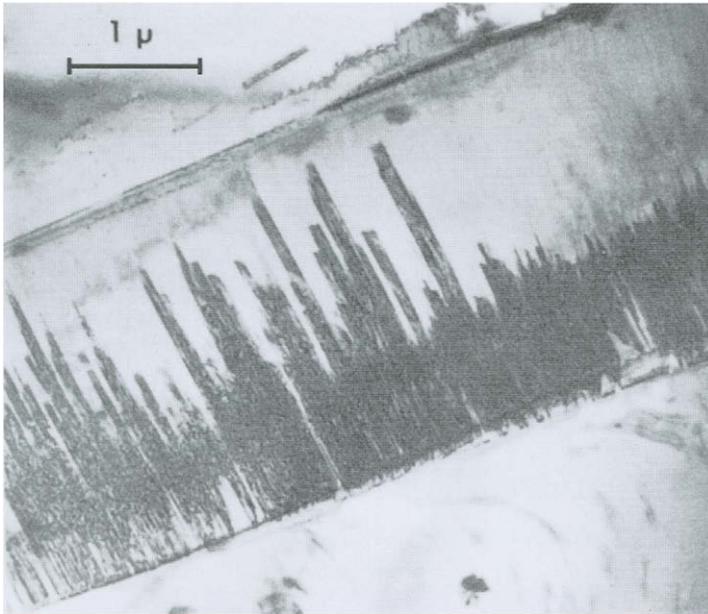


FIG. 22.35. Twins on one side only of a  $\{225\}$  plate in an iron-nickel-carbon alloy.

available to replace the old, and rather arbitrary, pair potentials. Electron microscopy has also made great strides and individual columns of atoms may be resolved when they are along low index directions and aligned parallel to the electron beam. Both methods require rational interfaces of fairly low indices, so that periodic boundary conditions can be applied in the simulations, and the high resolution images can be obtained from a close-packed direction. As pointed out in Chapter 20, the  $K_1$  planes of type I twins are particularly suitable for computer studies; such studies have recently been applied to twins and twinning dislocations in h.c.p. metals by Bacon and his collaborators (Serra and Bacon 1991, 1993; Serra *et al.*, 1988).

The coherent interface between f.c.c. and h.c.p. structures has been much studied by both methods. In a simulation of the  $(111)_c // (0001)_h$  interface in a material of ideal axial ratio in the h.c.p. phase, Sutton and Christian (1982) found, as anticipated, that the atomic structure of both phases was maintained up to the interface and that two atomic planes were common to the two structures. The structure of the conjugate invariant plane which is  $\{557\}_c$  for ideal axial ratio and irrational otherwise was studied with the assumption of ideal axial ratio. This structure was found to be relatively complex and motion of the interface would require a significant amount of atomic shuffling. Thus it is not surprising that the transformation always utilizes  $\{111\}$  as the habit plane, the more especially because the conjugate-invariant plane is irrational in all real materials.



FIG. 22.36. High resolution electron micrograph of fully coherent  $(111)_c// (001)_h$  interface in melt-spun cobalt.

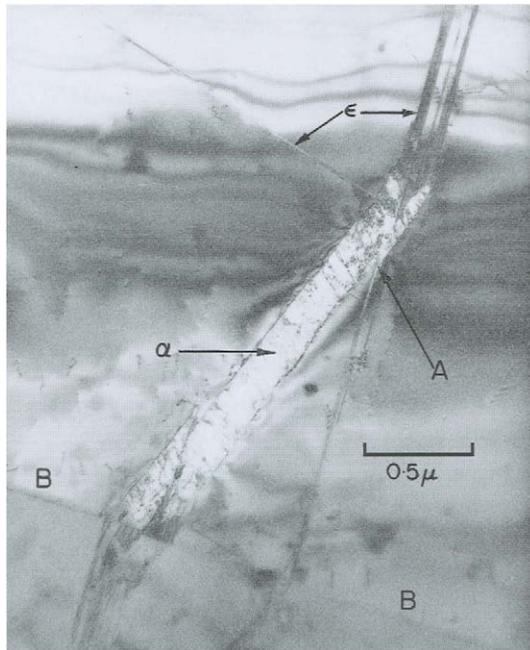


FIG. 22.37. Martensite forming from intersection of two martensite planes (courtesy of Venables, 1962).

There have been several investigations of the structure of the  $(111)_c// (0001)_h$  interface by high resolution electron microscopy. Figure 22.36 shows the interface in a rapidly solidified alloy; the occasional lattice planes which are highlighted represent stacking faults.

A f.c.c.–h.c.p. transformation also occurs in some iron–nickel–chromium alloys and related stainless steels of low stacking fault energy. This is called  $\varepsilon$ -martensite and is often an intermediate stage in the formation of  $\alpha$ -martensite in the same alloy. There is some evidence that, in these steels,  $\alpha$ -martensite forms from  $\varepsilon$ -martensite rather than directly from the austenite. Particularly favourable sites for the nucleation of  $\alpha$ -martensite appear to be the intersection of two  $\varepsilon$ -plates; an example is shown in Fig. 22.37 where a large (225) plate appears to have originated from the two  $\varepsilon$ -plates at the top of the micrograph. Such observations may be linked to the Boger–Burgers and Olson–Cohen models of martensite nucleation (see p. 1068).

#### REFERENCES

- AHLERS, M. (1974) *Z. Metallkd.* **65**, 636; (1975) *Metall. Trans.* **7**, 633.  
 BAIN, E. C. (1924) *Trans. Metall. Soc. AIME* **70**, 25.  
 BASINSKI, A. S. and CHRISTIAN, J. W. (1951–2) *J. Inst. Met.* **80**, 659; (1954a) *Acta Metall.* **2**, 101; (1954b) *Ibid.* **2**, 148; (1956) *Ibid.* **4**, 371.  
 BILBY, B. A. (1955) *Defects in Crystalline Solids*, p. 123, The Physical Society, London; (1956) *The Mechanism of Phase Transformations in Metals*, p. 334, The Institute of Metals, London.  
 BILBY, B. A. and CHRISTIAN, J. W. (1956) *The Mechanism of Phase Transformations in Metals*, p. 121, The Institute of Metals, London.  
 BOLLMAN, W. (1970) *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin.  
 BONNET, R. (1981a) *Acta Metall.* **29**, 437; (1981b) *Philos. Mag.* **43A**, 1195; (1981c) *Ibid.* **44**, 625; (1982) *J. Phys. (Paris)* **43**, 636; (1983) *Philos. Mag.* **51A**, 51.  
 BOWLES, J. S. (1951a) *Acta Crystallogr.* **4**, 162.  
 BOWLES, J. S., BARRETT, C. S., and GUTTMAN, L. (1950) *Trans. Metall. Soc. AIME* **188**, 1478.  
 BOWLES, J. S. and DUNNE, D. P. (1961) *Acta Metall.* **17**, 677.  
 BOWLES, J. S. and MACKENZIE, K. K. (1954a) *Acta Metall.* **2**, 129; (1954b) *Ibid.* **2**, 224.  
 BOWLES, J. S. and MALIN, A. S. (1960) *J. Aust. Inst. Met.* **5**, 131.  
 BOWLES, J. S. and SMITH, R. (1960) *Acta Metall.* **8**, 405.  
 BREDIS, J. F. and WAYMAN, C. M. (1962) *Trans. Metall. Soc. AIME* **224**, 1128.  
 BULLOUGH, R. and BILBY, R. A. (1956) *Proc. Phys. Soc. London, Ser. B* **69**, 1276.  
 BURGERS, W. G. (1934) *Physica* **1**, 561.  
 BUTCHER, B. R. and ROWE, A. H. (1956) *The Mechanism of Phase Transformations in Metals*, p. 229, The Institute of Metals, London.  
 CHRISTIAN, J. W. (1955–6) *J. Inst. Met.* **84**, 349; (1956) *The Mechanism of Phase Transformations in Metals*, p. 335, The Institute of Metals, London; (1975) *J. Phys. (Paris), Colloq.* **35-C7**, 65.  
 CROCKER, A. G. (1959) *Ph.D. Thesis*, Sheffield University, Sheffield, UK; (1962) *Acta Metall.* **10**, 113.  
 CROCKER, A. G. and BILBY, B. A. (1961a) *Acta Metall.* **9**, 678; (1961b) *Ibid.* **9**, 992.  
 DUNNE, D. P. and BOWLES, J. S. (1969) *Acta Metall.* **17**, 201.  
 EFSIC, E. J. and WAYMAN, C. M. (1967) *Trans. Metall. Soc. AIME* **239**, 873.  
 ENTWISLE, A. R. (1956) *The Mechanism of Phase Transformations in Metals*, p. 315, The Institute of Metals, London.  
 FRANK, F. C. (1953) *Acta Metall.* **1**, 15.

- GAUNT, P. and CHRISTIAN, J. W. (1959) *Acta Metall.* **7**, 534.
- GRENINGER, A. B. (1939) *Trans. Metall. Soc. AIME* **133**, 204.
- GRENINGER, A. B. and TROIANO, A. R. (1940) *Trans. Metall. Soc. AIME* **140**, 307; (1949) *Ibid.* **185**, 590.
- HITZENBERGER, C. and KARNTHALER, H. P. (1991) *Philos. Mag.* **64A**, 121.
- ISAITSHEV, I., KAMINSKY, E., and KURDJUMOV, G. V. (1938) *Trans. Metall. Soc. AIME* **128**, 361.
- KELLY, P. M. (1992a) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-92)*, eds. C. M. Wayman and J. Perkins, p. 293, Monterey Institute for Advanced Studies, Monterey, CA; (1992b) *Ibid.* p. 299.
- KELLY, P. M., BALL, C. J., and BLAKE, R. G. (1982) *Pap. 35th Annu. Conf. Aust. Inst. Met.*, pp. 97-99.
- KELLY, P. M. and NUTTING, J. (1960-1) *Proc. R. Soc. London, Ser. A* **259**, 45; (1961) *J. Iron Steel Inst.* **197**, 199.
- KRAUKLIS, L. P. and BOWLES, J. S. (1969) *Acta Metall.* **17**, 997.
- LIEBERMAN, D. S., WECHSLER, M. S., and READ, T. A. (1955) *J. Appl. Phys.* **26**, 473.
- LIEBERMAN, D. S., READ, T. A., and WECHSLER, M. S. (1957) *J. Appl. Phys.* **28**, 532.
- LIEBERMAN, D. S. (1969) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 167, The Institute of Metals, London.
- LIU, Y. C. (1956) *Trans. Metall. Soc. AIME* **206**, 1036.
- LIU, Y. C. and MARGOLIN, H. (1953) *Trans. Metall. Soc. AIME* **197**, 667.
- LOMER, W. M. (1956a) *The Mechanism of Phase Transformations in Metals*, p. 243, The Institute of Metals, London; (1956b) *Ibid.* p. 337.
- MACHLIN, E. S. and COHEN, M. (1951) *Trans. Metall. Soc. AIME* **191**, 746.
- MCDUGALL, P. G. and WAYMAN, C. M. (1992) *Martensite*, eds. G. B. Olson and W. S. Owen, p. 59, MIT Press, Cambridge, MA.
- MACKENZIE, J. K. and BOWLES, J. S. (1957) *Acta Metall.* **5**, 137.
- MUDDLE, B. C., KRAUKLIS, P., and BOWLES, J. S. (1976) *Acta Metall.* **24**, 371.
- NAKANISHI, N. (1961) *Trans. Jpn. Inst. Met.* **2**, 79.
- NISHIYAMA, Z. and SHIMIZU, K. (1962) *Mem. Inst. Sci. Ind. Res., Osaka Univ.* **19**, 213.
- NYE, J. F. (1953) *Acta Metall.* **1**, 153.
- OLSON, G. B. and COHEN, M. (1976) *Metall. Trans.* **7A**, 1897, 1905, 1915.
- OTSUKA, K. (1987) *Jpn. Inst. Met.* **35**.
- OTTE, H. M. (1968) *Scr. Metall.* **2**, 365.
- OTTE, H. M. and READ, T. A. (1957) *Trans. Metall. Soc. AIME* **209**, 412.
- PATTERSON, R. L. and WAYMAN, C. M. (1966) *Acta Metall.* **14**, 347.
- PETCH, N. J. (1943) *J. Iron Steel Inst.* **147**, 221.
- PIESCH, W. (1958-9) *J. Inst. Met.* **87**, 444.
- RAO, B. V. N. and THOMAS, G. (1979) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-s79)*, p. 122, MIT Press, Cambridge, MA.
- ROSS, N. H. D. and CROCKER, A. G. (1970) *Acta Metall.* **18**, 245.
- SATO, H., TOTI, R. S., and HONJO, G. (1967) *Acta Metall.* **15**, 1381.
- SERRA, A. and BACON, D. J. (1991) *Philos. Mag.* **63A**, 100; (1993) *Mater. Sci. Forum* **69**, 26.
- SERRA, A., BACON, D. J., and POND, R. C. (1988) *Acta Metall.* **36**, 3183.
- SANDVIK, B. P. and WAYMAN, C. M. (1983) *Metall. Trans.* **14**, 809, 823.
- SHIMIZU, K. (1962) *J. Phys. Soc. Jpn.* **17**, 508.
- SHIMIZU, K. and TANAKA, Y. (1978) *Trans. Jpn. Inst. Met.* **19**, 685.
- SUTTON, A. P. and CHRISTIAN, J. W. (1982) *J. Phys. (Paris), Colloq.* **43-C4**(Suppl. 12), 185.
- SWANN, P. R. and WARLIMONT, H. (1963) *Acta Metall.* **11**, 511.
- VENABLES, J. A. (1962) *Philos. Mag.* **7**, 35.
- WAITZ, T. and KARNTHALER, H. P. (1997) *Acta Metall.* **45**, 837.
- WARLIMONT, H. (1962) *Trans. Metall. Soc. AIME* **221**, 1270.
- WARLIMONT, H. and DELAËY, L. (1974) *Prog. Mater. Sci.* **18**.
- WATANABE, M. and WAYMAN, C. M. (1970a) *Scr. Metall.* **3**, 109; (1970b) *Metall. Trans.* **2**, 2229.
- WATANABE, M. and WAYMAN, C. M. (1971) *Metall. Trans.* **2**, 2221, 2229.
- WAYMAN, C. M. and MUDDLE, B. C. (1970) *Acta Metall.* **17**, 917.

- WAYMAN, C. M., HANAFEE, J. E., and READ, T. A. (1961) *Acta Metall.* **9**, 391.  
WECHSLER, M. S. (1959) *Acta Metall.* **7**, 793.  
WECHSLER, M. S., LIEBERMAN, D. S., and READ, T. A. (1953) *Trans. Metall. Soc. AIME* **197**, 1503.  
WECHSLER, M. S. and OTTE, H. M. (1961) *Acta Metall.* **9**, 117.  
WEINIG, S. and MACHLIN, E. S. (1954) *Trans. Metall. Soc. AIME* **200**, 1280.  
WILKENS, M. and WARLIMONT, H. (1963) *Acta Metall.* **11**, 1099.  
WILLIAMS, A. J., CAHN, R. W., and BARRETT, C. S. (1954) *Acta Metall.* **2**, 117.

## CHAPTER 23

# *Kinetics of Martensitic Transformations*

### 97. THE NUCLEATION OF MARTENSITE

As emphasized in Section 86, the overall kinetics of martensitic transformations are dependent mainly on the nucleation process as each plate grows rapidly to its limiting size. Spontaneous transformation is not possible on cooling until the temperature decreases below the  $T_0$  temperature at which the parent and product phases have equal free energies. In some alloys, the  $M_s$  temperature is only a few degrees below this thermodynamic transition temperature, but in steels and some other alloys a large driving force is needed to initiate the transformation. The amount of supercooling is small only when the shape deformation is small and this indicates that non-chemical terms in the overall energy change largely influence the temperature at which the transformation begins. Various thermodynamical estimates show that, in iron-carbon alloys, the driving force at the  $M_s$  temperature is approximately constant and is about  $1200 \text{ J mole}^{-1}$ . In iron-nickel and iron-chromium alloys, calculations show very much lower driving forces which increase rapidly with increasing alloy content. This was formerly rather puzzling but, as noted on p. 710, it now appears probable that at low alloy contents a massive transformation rather than a martensitic transformation is obtained in rapidly cooled alloys.

The calculation of the driving force for  $\gamma$  loop elements like chromium or molybdenum is rather complex. The existence of a  $\gamma$  loop implies that the equilibrium temperature between the phases is raised above that for pure iron, but at the same time these elements are observed to depress the  $M_s$  temperature. One explanation of the apparent anomaly is given by Zener's division of the free energy change in iron into magnetic and non-magnetic terms (see p. 109) and consideration of the separate effect of the alloying element on each of these. Further details of these thermodynamical calculations would be out of place here; the reader is referred to Kaufman and Cohen (1956), Owen and Gilbert (1960) and Kaufman and Hillert (1992). The kinetics of martensite formation, especially isothermally, have been comprehensively reviewed by Raghavan (1992).

An isolated martensite plate forming with a large shape deformation will produce an appreciable elastic strain energy as discussed in Section 52. For a given shape, this positive energy term will be proportional to the volume of the plate and must be added to the negative energy term resulting from the phase change. Other non-chemical terms to be included in the overall energy balance include the energy of the interfaces, energy stored in

the parent and/or the product phases as dislocations and point defects by plastic deformation caused by internal stresses and kinetic energy dissipation by the moving interface, moving dislocations, etc. Transformation will only be possible if the overall energy balance is favourable; i.e. if the negative chemical term has a magnitude greater than all the other positive energies. Because lattice deformation and kinetic energies act so as to reduce the strain energy, a calculation based on the assumption that the shape strain is accommodated entirely elastically will give an upper limit to the chemical driving force required. However, in addition to all these energy terms, there may be an appreciable activation energy for the nucleation of the martensite which will cause further supercooling.

Nucleation is one kinetic factor which affects the temperature at which transformation can begin. Another restriction which need not be considered in most other transformations is that a martensite plate may form too rapidly for the released latent heat to escape, and the local temperature then begins to rise rapidly. Krisement *et al.* (1954) pointed out that the rapid formation of a martensite plate approximates to an adiabatic rather than an isothermal process, and they proposed that  $M_s$  must be sufficiently low in temperature that the local temperature cannot rise above the strain energy-modified equilibrium temperature  $T_0''$ . Their calculations indicated that this effect could be quite large in some iron-carbon alloys; they estimated the non-chemical terms in the driving force to be about 275–335 J mole<sup>-1</sup> and attributed the remainder of the supercooling to the adiabatic heating. The calculations were improved by Hillert (1958), who showed that the minimum temperature hysteresis between cooling and heating transformations on the assumption that the transformation takes place adiabatically and reversibly is given by

$$T_1 - T_0'' = -T^{\Delta M} \Delta s^{\Delta M} / c_p \quad (97.1)$$

where  $T^{\Delta M}$  is the equilibrium temperature,  $\Delta s^{\Delta M}$  is the entropy change on formation of martensite and  $c_p$  is the specific heat.  $T_0''$  is the modified equilibrium temperature, including both strain energy and adiabatic heating.

The temperature  $T_0''$  is the highest temperature at which formation of martensite will not raise the local temperature momentarily above  $T^{\Delta M}$ .  $T_1$  is similarly the minimum temperature which will ensure that the local temperature will not fall below  $T^{\Delta M}$ . The actual hysteresis may be appreciably larger than that given by eqn. (97.1) because the non-chemical elastic energy must be included on the basis that it is irreversible, so that it both lowers  $M_s$  and raises  $T_0''$ . According to the adiabatic hypothesis, eqn. (97.1) gives the minimum hysteresis that can be expected in a transformation, even with zero strain energy, and might correspond to  $A_s - M_s$ . With reasonable values of  $T^{\Delta M} = 450$  K,  $-\Delta s^{\Delta M} = 6$  J K<sup>-1</sup> mole<sup>-1</sup> and  $c_p = 30$  J K<sup>-1</sup> mole<sup>-1</sup>, Hillert obtained good agreement with the results of Kaufman and Cohen shown in Fig. 21.6. This is a much smaller effect than that calculated by Krisement *et al.* and it implies that most of the hysteresis is not due to the adiabatic requirement. In addition, if the hysteresis is due to adiabatic conditions, it is difficult to understand the effect of stress on the transformation which, under some circumstances, can reduce the hysteresis virtually to zero (see Fig. 21.5).

The other possible kinetic factors which lead to a low temperature for initiation of the transformation arise from nucleation of the new phase. Theories of nucleation suffer from familiar limitations imposed by the lack of knowledge of what constitutes a critical stage in the growth of a very small plate, and the semi-coherent martensite interfaces with their arrays of dislocations or twins merely exacerbate the problem. It is not clear whether the critical nucleus should be a fully formed plate with this structure or only a fully coherent region with the correct lattice correspondence. The first alternative seems more probable as the nucleus or embryo will not be able to grow until the habit plane has been adjusted to a macroscopic invariant plane strain. There is good evidence (Patel and Cohen, 1953; Fisher and Turnbull, 1953) that the effects of stress on the transformation are correctly predicted by supposing the stress to interact with the strains of the shape deformation, rather than with those of the lattice deformation. As already noted, the change in  $M_s$  produced by a given stress system seems to be correctly predicted by postulating a fixed value for the net driving force (chemical plus mechanical) at the  $M_s$  temperature. If nucleation in the operational sense (not necessarily classic nucleation; see below) determines the  $M_s$  temperature, this implies that the critical nucleus condition must be one in which the total shape change is already established.

There are three main kinds of nucleation to discuss. The first of these is classical nucleation by thermal fluctuations, either homogeneously or heterogeneously at suitable defects in the solid, or at grain boundaries. The second models are structural; a critical configuration is assumed to form or exist in the parent with the structure of the required nucleus or of some critical configuration on the way to the eventual nucleus. The third possibility is non-classical; a configuration representing a martensite region is assumed to arise spontaneously as a result of a lattice instability; something like a strain spinodal has to be postulated and some lattice elastic constants tend to zero as  $M_s$  is approached.

Some displacive phase transitions involve very small principal strains and are found to be quasi-continuous; they result from a lattice instability known as a soft phonon mode. If an elastic stiffness constant tends to zero as the transformation range is approached, the instability appears as a static wave. In martensitic transformations, there is often some softening immediately above  $M_s$  and the stiffness, contrary to the usual behaviour, decreases with decreasing temperature. However, such an effect is not observed in steels and, even in copper alloys, the appropriate shear stiffness extrapolates to a finite value at  $M_s$ .

Pursuing this type of theory a little further, it is usual to write the free energy of the assembly as a Landau expansion in a generalized order parameter which is assumed to be the strain along some critical path leading from one structure to the other. In considering lattice stability against a long-range fluctuation in stress, it seems probable that a strain gradient term should be introduced by analogy with the composition gradient of spinodal decomposition by diffusion. As a generalized order parameter, Olson and Roitburd (1992) use the quantity  $\eta = \varepsilon/\varepsilon_f$ , where  $\varepsilon$  represents the strain along some path and  $\varepsilon_f$  represents the total strain, i.e. the martensitic structure. Thus  $\eta$  varies from zero to one and the first question which arises is whether the nucleus is represented by the condition  $\eta = 1$  or whether  $\eta < 1$  at the nucleation stage.

In the classical theory of homogeneous nucleation, the nucleus is usually treated as an oblate spheroid of axes  $R$ ,  $R$  and  $y$ , and with the final structure of the martensite (i.e.  $\eta = 1$ ). The treatment follows that given for homogeneous nucleation in Chapter 20 (pp. 908–950), with the negative free energy change produced by the phase transformation replacing the mechanical work done by the external stress.

In one of the first theories of martensite nucleation, Cohen *et al.* (1950) introduced the idea of the “reaction path”, a progressively increasing deformation of the parent structure which leads ultimately to the martensitic product structure. They considered the interaction of a distorted rectangular volume with a dislocation array and assumed that the free energy along the reaction path, although generally decreasing, would have a local maximum value at some intermediate configuration, the overcoming of which would represent the critical condition for nucleation. The maximum vanishes at some critical value of the driving force, and the parent lattice is then mechanically unstable; alternatively the maximum may be overcome with the aid of thermal fluctuations giving isothermal nucleation (Machlin and Cohen, 1952).

At about the same time as the reaction path theory was developed, Kurdjumov (1949) and Fisher *et al.* (1949) were applying classical nucleation theory to martensite. The main assumption of this theory is that the critical nucleus is a fully formed region of martensite, in contrast to the assumption of Cohen *et al.* that it is an configuration intermediate between the parent and product phases, and these two possibilities have dominated much of the subsequent discussion.

This gives the critical radius and free energy as

$$R_c = 4\sigma Z / (\Delta'g^{\text{AM}})^2 \quad (97.2)$$

$$\Delta G_c = 32\pi\sigma^3 Z^2 / 3(\Delta'g^{\text{AM}})^4 \quad (97.3)$$

and the energy  $\Delta G_c$  is much too large to be overcome by thermal fluctuations. Hollomon and Turnbull (1953) estimated that if the energy of the interface were as low as  $10 \text{ mJ m}^{-2}$ , the nucleation rate would be large enough to explain their experimental results shown in Fig. 21.5. Unfortunately, however, estimates of the interfacial free energy of a plate with a glissile interface like that of Fig. 22.28 are in the region of  $200 \text{ mJ m}^{-2}$  and this gives a negligible nucleation rate. Fisher and Turnbull (1953) produced a modification of the theory in which they used a strain-dependent interfacial energy, treating the martensite as linear elastic, and in this way they reduced the activation free energy for nucleation to about one-half of its above value.

Knapp and Dehlinger (1956) also treated the nucleus as an oblate spheroid and, with a surface free energy of  $200 \text{ mJ m}^{-2}$ , they obtained similar estimates of the nucleus size at the maximum of the free energy curve. However, they assumed a critical growth condition, namely that the nucleus will grow rapidly and spontaneously as soon as the overall change in free energy is negative. This condition

$$\begin{aligned} R &= 25\sigma Z / 4(\Delta'g^{\text{AM}})^2 \\ \Delta G &= 0 \end{aligned} \quad (97.4)$$

is not attained until the nucleus size is slightly larger than that of the classical critical nucleus, so that the nucleation rate predicted by this theory is again quite negligible. However, the theory is of interest in drawing attention to the possibility that the nucleation rate may be determined not by the energetic considerations alone but by a condition governing the onset of rapid growth. This can give an effective or operational nucleation rate.

The onset of rapid growth may be related to the size of nucleus needed to enforce spontaneous nucleation of dislocation loops (or, more realistically, of twins) at the edge of the growing plate (see p. 1050). Kaufman and Cohen discussed this possibility in terms of a very approximate model, and suggested that the critical radius of such a nucleus is about 36/25 times larger than that given by eqn. (97.2).

Kaufman and Cohen (1958) adapted the classical model to heterogeneous nucleation through interaction with a dislocation array leading to an effectively negative interfacial free energy. Above the  $M_s$  (or rather the  $T_0$ ) temperature where the bulk martensite phase is not thermodynamically stable, this interaction would lead to the local stabilization of martensitic embryos, and curves of free energy versus size would then show local minima. Kaufman and Cohen assumed that there would be a distribution of such embryos of varying sizes, and that the  $M_s$  temperature corresponds to the largest embryos becoming supercritical on cooling. As cooling continues, more and more of the existing embryos will exceed the critical nucleus size and will grow into a product plate, thus giving athermal characteristics. Isothermal characteristics result when the pre-existing embryos at the transformation temperature are all smaller than the critical size but can attain this size by thermal fluctuations. Unfortunately, however, there is no experimental evidence to support this ingenious theory and attempts to detect pre-existing embryos have been unsuccessful.

The above discussion applies essentially to nucleation at relatively small driving forces. At large driving forces, nucleation occurs close to the point where the lattice becomes unstable, and this tends to  $\eta=0$  at very large driving forces. In most practical situations this condition is inaccessible because heterogeneous nucleation will have intervened, but some experiments by Liu *et al.* (1992) have apparently detected homogeneous nucleation in iron-cobalt alloys. They used the small droplet technique, the iron-cobalt particles being dispersed in a matrix of copper. The measured  $M_s$  temperatures of the particles corresponded to a driving force of about 10 kJ mole<sup>-1</sup>, a factor of seven higher than the known driving force for the bulk alloy of the same composition. Liu *et al.* argue that this must represent homogeneous nucleation of fully coherent particles in the shape of an oblate spheroid or pillbox, semi-coherent particles being ruled out because of the high activation energy derived from the high value of the interfacial energy.

Knapp and Dehlinger (1956) developed a theory of nucleation which combines the formalism of the classic approach with a calculation of interfacial energy based on Frank's model of the interface. As already mentioned, instead of using the criterion  $\partial\Delta G/\partial R = \partial\Delta G/\partial y = 0$  for the saddle point, they postulate that the nucleus will grow rapidly and spontaneously as soon as  $\Delta G$  becomes negative.

Dividing the expression for  $\Delta G$  [see eqn. (97.3)] by the volume of the nucleus gives

$$\Delta G/V = \Delta' y^{\text{AM}}/v^{\text{M}} + 3\sigma/2y + Zy/R \quad (97.5)$$

The first term in this equation is the net change in chemical free energy which provides the driving force for the reaction, the third term gives the effects of strain energy and the second term represents the interfacial free energy. Knapp and Dehlinger's treatment gives the same relation between  $R$  and  $y$ , namely

$$y = -(\sigma R/Z)^{1/2} \quad (97.6)$$

if the plates are assumed to be spheroids, and differs from this only by a factor  $(4/3)^{1/2}$  if they are taken as lenticular discs as in eqn. (97.5). Knapp and Dehlinger's condition that the net driving force be positive does not seem sensible, but there may be other factors which limit cataclysmic growth until some situation beyond the saddle point is attained. This means that the growth condition is that the embryo reaches a critical size larger than that which corresponds to a saddle point on the energy surface. This assumption [see eqns. (97.4)] gives the value of  $Z$  derived by Knapp and Dehlinger, which is slightly smaller than that given in Chapter 10 although the difference is not significant. The value of the energy of the martensite interface is estimated by Knapp and Dehlinger and by Kaufman and Cohen on the basis of Frank's model of the interface as about  $200 \text{ mJ m}^{-2}$ . With this value, the critical size at which a nucleus can grow freely is very large and the probability of such a nucleus forming by spontaneous thermal fluctuations is zero.

Homogeneous nucleation theory at high driving forces deviates from the classical theory and resembles the Cahn-Hilliard theory for nucleation just outside the spinodal. In the martensite case, as a point of lattice instability is approached, the critical nucleus size rises and the free energy of formation falls, both abruptly. The interface becomes diffuse and the surface free energy has to be replaced by a gradient energy term; the consequent behaviour is described by an equation of the Cahn-Hilliard type (Olson and Cohen, 1982). Whether or not the experiments on nucleation in iron-cobalt droplets reached this situation; or whether they represented an intermediate situation with  $\eta < 1$ , the treatment rather like the classical equations (Fisher and Turnbull, 1953), is not clear. It is, however, obvious that in most bulk specimens nucleation is heterogeneous and takes place at driving forces much smaller than those which approach the stability limit. The critical driving force for lattice instability is estimated as  $200\text{--}450 \text{ J mole}^{-1}$  for sodium but as  $14 \text{ kJ mole}^{-1}$  for iron (Olson and Roitburd, 1992). Zhang and Li (1992) have investigated the sites at which various forms of ferrous martensite appear. They confirm that martensite forms from four or five overlapping stacking faults, as in the Olson-Cohen model, and in cobalt alloys, in broad agreement with the results of Karnthaler and his co-workers (Hitzenberger and Karnthaler, 1991; Waitz and Karnthaler, 1997) and of Howe *et al.* (1985, 1987). Twin and grain boundaries, especially at triple points, were found to be favourable sites for lenticular and thin-plate martensites, but not for butterfly martensite which formed inside austenite grains. It is suggested that lenticular martensite forms because emissary dislocations from the twins run forward and provide the lattice-invariant deformation. Tsuzuki *et al.* (1995) have reported experiments on the relative

potencies of different grain boundaries as nucleation catalysts. They found that single crystals of an Fe–32 wt.% Ni alloy had an  $M_s$  temperature of 88 K whilst polycrystalline specimens had an  $M_s$  of 154 K. Bicrystal specimens containing either twin boundaries or random high angle boundaries gave the same  $M_s$  as the single crystals. The results are interpreted to mean that only certain special boundaries which are neither high angle nor twin are potent nucleation sites. However, if there are very few such boundaries (their minimum estimate is at least one in 200), it is difficult to see that much transformation would be provided in polycrystalline specimens given that martensite plates cannot cross grain boundaries.

Among the specific structural models for nucleation, those of Kaufman and Cohen (1958) (discussed above) and Olson and Cohen (1979) should be mentioned. In the model of the oblate spheroidal nucleus with dislocations to satisfy the invariant plane strain macroscopic condition, they came up against the difficulty already mentioned, namely that with the best estimates of the interfacial free energy available, the activation energy for formation of such a nucleus homogeneously is much too high. The Olson–Cohen model is based on a ribbon-shaped nucleus formed from a wall of dislocations which dissociate to give stacking faults at regular spacings one above the other. In the case of the f.c.c.–h.c.p. transformations, this model is almost certainly correct, and it is confirmed by the in-situ observations, already summarized, that the product phase forms first as a thin layer parallel to the close-packed  $\{111\}$  planes. Note that this model gives apparently barrierless nucleation (although there may well be a barrier against thickening of the initial six to 12 layer plate). Olson and Cohen also regarded the formation of such a group of stacking faults (one every third layer) as the first stage in the barrierless formation of a b.c.c. structure from a f.c.c. structure, the further adjustments being supposed to take place spontaneously. This theory is a version of the Bogers Burgers mechanism for the crystallographic change, and the authors attributed thermal nucleation to the thermally activated overcoming of an interface friction stress.

Whatever the merits of this particular proposal, it is difficult not to accept that under some conditions martensitic nucleation may occur spontaneously without the intervention of a barrier which has to be overcome by thermal fluctuations. The experiments of Kulin and Cohen (1950) showed conclusively that martensite may be formed under the influence of either temperature or an external stress at temperatures as low as 4.2 K. Barrierless nucleation can presumably result from defect interactions of unknown types, of which the Olson–Cohen model is one possible example, and the nucleation rate is then controlled by interface mobility.

Nucleation without a barrier was considered by Suezawa and Cook (1980), who considered the interaction of a circular plate with a linear defect, and by Roitburd (1981), who considered a rectangular nucleus. With a barrierless nucleation, the kinetics of nucleation might be due to interface kinetics, i.e. the lattice resistance or some other obstacle to interface migration, as first proposed by Kaufman and Cohen (1958).

In a comprehensive review of the current theories, Olson and Roitburd (1992) re-emphasize that strongly non-classical behaviour can only result at extreme driving forces, very close to the point of lattice instability. Such a condition is unlikely ever to arise

in bulk specimens because of the intervention of heterogeneous nucleation by a classical path. However, homogeneous nucleation has probably been achieved in the experiments of Liu *et al.* on small droplets of iron-cobalt in copper.

A notable feature of the nucleation of martensite is that the first crystals to form trigger off many other crystals, a phenomenon which is known as "autocatalytic" or "sympathetic" nucleation. In extreme cases, probably when the potency of some autocatalytic defects is greater than that of most of the defects causing initial nucleation, the first trigger will initiate a complex chain of transformations which is referred to as a "burst". Autocatalytic nucleation is thought to be due to the effects of the internal stress field generated by the first nucleus.

### 98. FORMAL KINETICS OF MARTENSITE FORMATION

In this section, the factors which determine the volume fraction of martensite formed from the parent phase as a function of time in an isothermal reaction, or as a function of temperature in an athermal reaction, are discussed. Thermoelastic martensites need not be considered, as the previous description shows that they are always in equilibrium with the matrix, and there is no kinetic problem. In many martensitic transformations, however, a plate forms in a very short time, after which it ceases to grow, and further transformation has to be initiated by the formation of new plates. In most cases, edgewise growth is limited by major obstacles such as grain boundaries, other martensite plates, inclusions, etc., and one effect of deforming the specimen before transformation (as in ausforming) is to reduce the size of the martensite plates by increasing the initial density of such obstacles. However, an additional mechanism is required to explain why the plate does not continue to thicken after edgewise growth has ceased. In thermoelastic martensite such thickening leads to elastic equilibrium between plate and matrix; it seems certain that, in other cases, plastic deformation begins before such equilibrium is attained. This deformation will make it more difficult for the interface to advance into the matrix and eventually it will become jammed either by sweeping up defects or because the deformation destroys the coherency between the martensite and the matrix.

Experimental evidence in support of this concept is provided by the different behaviour of martensite with small shape deformations, or self-accommodating plate systems, and also by thin film electron microscopy. Dislocations are not visible in densely twinned martensite plates, but very dense tangles of dislocations are observed in the matrix adjacent to lenticular plates as shown in Fig. 23.1. It is often found that the plates of martensite formed in any one specimen have a constant length:thickness ratio and this is consistent with the hypothesis that growth ceases because of plastic deformation. Two kinds of reaction curve have to be considered in the formal kinetics of martensite formation, namely fraction transformed against time at constant temperature for isothermal martensite and fraction transformed against temperature for athermal martensite. The isothermal curves show that the reaction rate decreases rapidly during the course of transformation. If the operational nucleation rate is assumed to be constant, which seems reasonable at constant driving force, the reduction in reaction rate must be



FIG. 23.1. Dense dislocation tangles in retained austenite around twinned martensite plates.

caused by a progressive change in the volume of martensite formed from each nucleus. A natural explanation of the effect is simply that the martensite plates constantly subdivide the parent matrix, so that later plates form in smaller parent phase volumes. A formal theory of this effect was given by Fisher (1953a).

Let each plate subdivide the matrix so that after  $n$  plates have formed the next plate is confined to a region comprising  $1/n$  of the remaining untransformed volume. If each plate transforms a constant fraction  $f$  of the parent volume in which it forms, the volume fraction  $\zeta$  which has transformed to martensite is related to the number of plates formed by

$$\begin{aligned} d\zeta/dn &= f(1 - \zeta)/n \\ f \ln(n/n_0) &= -\ln(1 - \zeta) \end{aligned} \quad (98.1)$$

where  $n_0$  is the extent to which the parent phase is subdivided at the beginning of transformation ( $\zeta = 0$ ), and is essentially the number of parent phase grains, or perhaps subgrains, in the specimen. The effective number of subdivisions is the sum of  $n_0$  and the number of plates which have formed since the beginning of the transformation.

Suppose there is a constant nucleation rate  ${}^vI$  per unit volume. Then the initial overall nucleation rate is  $I = {}^vIV$  and the nucleation rate after some transformation is  $I(1 - \zeta)$ . Hence the total number of partitioning elements is given by

$$n = n_0 + \int f(1 - \zeta) dt \quad (98.2)$$

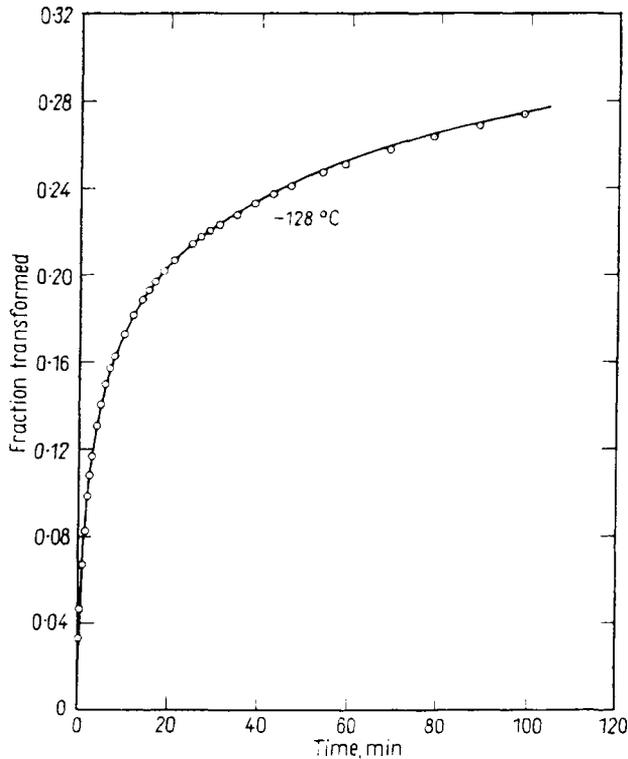


FIG. 23.2. Formation of isothermal martensite in an iron nickel manganese alloy (after Fisher, 1953). The experimental points are data of Cech and Hollomon (1953); the curve is calculated from eqn. (98.3).

If  $n$  is now eliminated between eqns. (98.1) and (98.2), the resulting equation can be solved for  $\zeta$  by differentiating, rearranging and integrating, and this gives

$$\zeta = 1 - [1 + (1 + f)It/n_0]^{-f/(1+f)} \quad (98.3)$$

Fisher fitted eqn. (98.3) to the isothermal transformation data of Cech and Hollomon (1953), treating  $f$  and  $I/n_0$  as disposable parameters. Very good agreement with experiment can be obtained in this way, as shown in Fig. 23.2. The values of  $f$  and  $I/n_0$  so determined are shown in Fig. 23.3 as a function of reaction temperature. The theory suggests that  $f$  should be a slowly varying function of temperature, the major change in reaction kinetics coming from the variation in  $I$ . The experimental values of  $f$  show a variation by a factor of five and have a maximum value which is not easily explained. However, there is a much larger variation in  $I$  by a factor of about 70 and the curve shows a sharp maximum at  $-128^\circ\text{C}$ . The results are thus generally consistent with both the partitioning assumption and with the classical theory of nucleation as outlined in the last section. No account has been taken in this theory of the autocatalytic effect of previously formed martensite; this has been allowed for in a kinetic theory developed by Raghavan and Entwisle (1980).

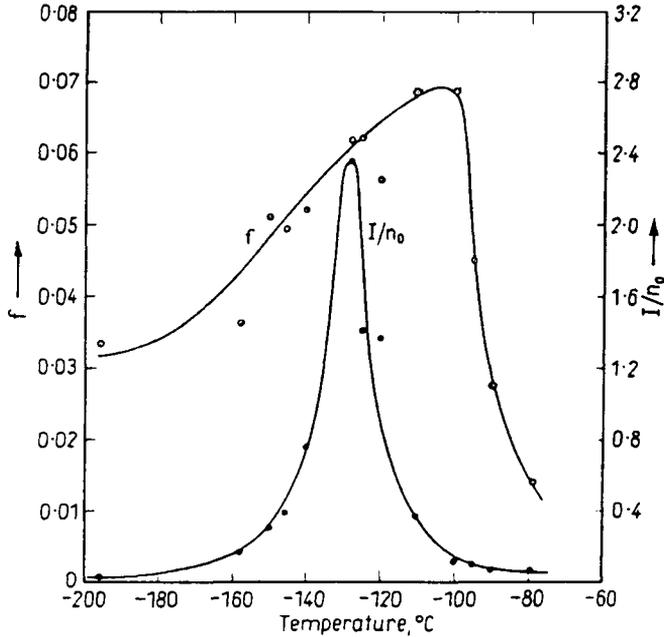


FIG. 23.3. Values of  $f$  and  $I/n_0$  in eqn. (98.3) as functions of temperature (after Fisher, 1953a). The values plotted are those which best fit the isothermal transformation data of Fig. 21.11.

The partitioning assumption can also be used to describe the variation of fraction transformed with temperature in an athermal transformation in terms of the variation with temperature of the number of nuclei formed per unit volume of untransformed parent phase. In eqn. (98.3) it is now replaced by  $n'$ , the number of nuclei which would have formed in an untransformed volume equal to the total volume of the specimen. (Clearly  $I$ ,  $n$ ,  $n_0$  and  $n'$  may alternatively be defined for unit volume instead of for the whole specimen as above.) Fisher (1953b) has used this modified equation to discuss athermal transformation in iron nickel alloys in terms of the assumption that the parent athermal characteristics are due to rapid thermal nucleation during cooling. The calculated width of the transformation range for up to 50% transformation, above which the equation is considered unreliable, is in moderately good agreement with the experimental results.

A number of empirical relations have been suggested for the variation of volume fraction transformed with temperature for athermal martensite. Koistinen and Marburger (1959) obtained accurate data on pure iron-carbon alloys with an X-ray diffractometer, and found that the results fitted a curve

$$1 - \zeta = \exp[-0.011(M_s - T)] \quad (98.4)$$

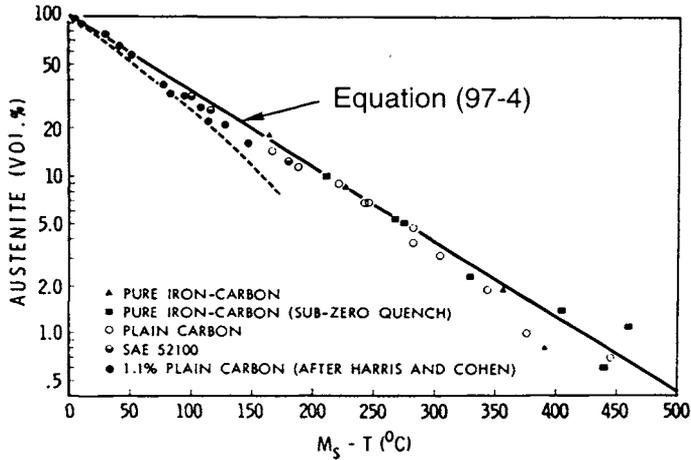


FIG. 23.4. Experimental results for volume fraction of austenite (and hence martensite) versus supercooling below  $M_s$ , compared with empirical equations.

where  $T$  is the reaction temperature. As shown in Fig. 23.4, this relation gives a reasonably good fit with experimental data for plain carbon steels, including some for which Harris and Cohen (1949) proposed a more elaborate relation. It will be seen that, according to the equation, complete transformation is never attained, i.e. there is no true  $M_f$  temperature.

As is implied by Fig. 23.4, the athermal transformation curves of many steels have a similar shape. A plot of  $1 - \zeta$  against temperature has an initial curved region and then leads into a relation which is linear between 5% and 60% martensite according to many workers. Beyond this linear range, the increment of transformation for each unit of cooling decreases and approaches zero towards the  $M_f$  temperature. Instead of attempting to describe the whole curve by a single equation, Brook *et al.* (1960) concentrated attention on the linear region, and they showed that the slope of the line (i.e. the increment in percentage transformed per degree Celsius) is directly proportional to the  $M_s$  temperature in both chromium and nickel steels, although the constant of proportionality is slightly different in the two cases. Thermodynamic calculations show that this result is an expression of a more fundamental relation, namely that the slope is proportional to the temperature derivative of the driving force  $\Delta g^{AM}/\partial T$ . There are some anomalies for  $M_s$  temperatures below 60°C, i.e. at high alloy and carbon contents.

Most experimental results show that the fraction of austenite converted to martensite is a function only of  $(M_s - T)$  and is unaffected by variations in composition or austenitizing treatment except insofar as these affect  $M_s$ . This means that knowledge of the  $M_s$  temperature should enable the volume fraction of martensite at any lower temperature to be predicted. The results of Brook *et al.* suggest that this is due to a fundamental correlation between the increment of transformation and the change in driving force with temperature.

Two previously described kinetic effects may intervene and modify the above simple considerations. One of these is the burst phenomenon, which produces a large amount of

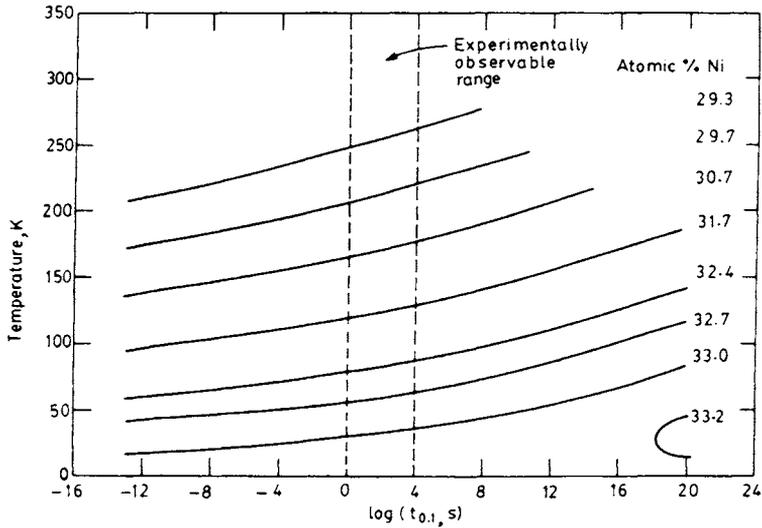


FIG. 23.5. Calculated time for the initial 0.1% martensitic transformation as a function of temperature and nickel content.

transformation in a small temperature interval. The other is stabilization, which inhibits transformation when cooling is resumed. Simple curves of percentage transformed versus temperature cannot be expected when the burst effect is significant; according to one interpretation of stabilization its main effect is to eliminate the autocatalytic formation of martensite which is the reason for bursts.

Raghavan (1992) has given a much more complete discussion of the kinetics of martensite formation. He emphasized the contrast between carbon-nickel manganese alloys which show C-curve behaviour and iron-nickel alloys which do not. Raghavan also discussed the possibility (Fisher, 1953) that athermal martensite is simply very fast thermal martensite with an activation energy which is very small. He cited experimental results which show that the activation energy is linearly proportional to the chemical free energy driving force, the constant of proportionality being much larger for iron-nickel alloys than for iron-nickel manganese alloys. By extrapolating to lower temperatures and higher nickel contents (a highly dubious procedure in view of the very small range over which measurements can be made), the results shown in Fig. 23.5 are obtained. These results suggest that it is impossible to distinguish experimentally between a true athermal process and a thermal process with a very small activation energy.

#### REFERENCES

- BROOK, R., ENTWISLE, A. R., and IBRAHIM, E. F. (1960) *J. Iron Steel Inst.* **195**, 292.  
 CECH, R. E. and HOLLOWOM, J. H. (1953) *Trans. Metall. Soc. AIME* **197**, 685.

- COHEN, M., MACHLIN, E. S., and PARANJPE, V. G. (1950) *Thermodynamics in Physical Metallurgy*, p. 242, American Society for Metals, Cleveland, OH.
- FISHER, J. C. (1953a) *Acta Metall.* **1**, 32; (1953b) *Trans. Metall. Soc. AIME* **187**, 918.
- FISHER, J. C., HOLLOMON, J. H., and TURNBULL, D. (1949) *Trans. Metall. Soc. AIME* **185**, 691.
- FISHER, J. C. and TURNBULL, D. (1953) *Acta Metall.* **1**, 310.
- HARRIS, W. T. and COHEN, M. (1949) *Trans. Metall. Soc. AIME* **180**, 447.
- HILLERT, M. (1958) *Acta Metall.* **6**, 122.
- HITZENBERGER, C. and KARNTHALER, H. P. (1991) *Philos. Mag.* **64A**, 151.
- HOLLOMON, J. H. and TURNBULL, D. (1953) *Prog. Met. Phys.* **4**, 333.
- HOWE, J. M., AARONSON, H. I., and GRONSKY, R. (1985) *Acta Metall.* **33**, 639; (1987) *Philos. Mag.* **56A**, 31.
- KAUFMAN, L. and COHEN, M. (1956) *The Mechanism of Phase Transformations in Metals*, p. 187, The Institute of Metals, London; (1958) *Prog. Met. Phys.* **7**, 165.
- KAUFMAN, L. and HILLERT, M. (1992) *Martensite*, eds. G. B. Olson and W. S. Owen, p. 41, MIT Press, Cambridge, MA.
- KNAPP, H. and DEHLINGER, U. (1956) *Acta Metall.* **4**, 289.
- KOISTINEN, D. P. and MARBURGER, R. E. (1959) *Acta Metall.* **7**, 59.
- KRISMENT, O., HOUDRIMENT, E., and WEVER, F. (1954) *Rev. Metall.* **51**, 401.
- KULIN, S. A. and COHEN, M. (1950) *Trans. Metall. Soc. AIME* **188**, 1138.
- KURDJUMOV, G. V. (1949) *Dokl. Akad. Nauk. SSSR* **60**, 1543.
- LIU, M., OLSON, G. B., and COHEN, M. (1992) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-92)*, eds. C. M. Wayman and J. Perkins, p. 191, Monterey Institute for Advanced Studies, Monterey, CA.
- MACHLIN, E. S. and COHEN, M. (1952) *Trans. Metall. Soc. AIME* **194**, 489.
- OLSON, G. B. and COHEN, M. (1979) *Acta Metall.* **27**, 1079; (1982) *J. Phys. (Paris), Colloq.* **43-C4**, 75.
- OLSON, G. B. and ROITBURD, A. L. (1992) *Martensite*, eds. G. B. Olson and W. S. Owen, p. 149, MIT Press, Cambridge, MA.
- OWEN, W. S. and GILBERT, A. (1960) *J. Iron Steel Inst.* **196**, 142.
- PATEL, J. R. and COHEN, M. (1953) *Acta Metall.* **1**, 531.
- RAGHAVAN, V. and ENTWISLE, A. R. (1980) *The Physical Properties of Martensite and Bainite*, p. 30, The Iron and Steel Institute, London.
- ROITBURD, A. L. (1981) *Solid State Physics*, Vol. 33, eds. H. Ehrenreich, F. Seitz and D. Turnbull, p. 317, Academic Press, New York.
- SUEZAWA, M. and COOK, H. E. (1980) *Acta Metall.* **28**, 423.
- TSUZUKI, K., HARADE, N., and MAKI, T. (1995) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-95) (J. de Phys. IV, Colloq.)*, eds. R. Gotthardt and J. Van Humbeeck, p. C8-167.
- WAITZ, T. and KARNTHALER, H. P. (1997) *Acta Mater.* **45**, 837.
- ZHANG, X. and LI, Y. Y. (1992) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-92)*, eds. C. M. Wayman and J. Perkins, p. 203, Monterey Institute for Advanced Studies, Monterey, CA.

## CHAPTER 24

# *Rapid Solidification*

### **99. METHODS FOR RAPID COOLING OF A LIQUID; GENERAL PROPERTIES OF RESULTANT SOLID**

Although the rapid cooling of high temperature solid structures by quenching into water or oil had been practised for centuries, it was not until 1960 that very rapid cooling of liquids from temperatures above the melting point of the solid phase was first achieved (Duwez *et al.*, 1960; Klement *et al.*, 1960). The method used by these authors was to bring a thin layer of liquid suddenly into contact with a solid of high thermal conductivity. Small droplets of the liquid were accelerated by a shock wave and allowed to impinge on to a copper target, thus producing flat discs of irregular shape. The significance of this experiment was that, if the cooling was initiated in the liquid state, much higher cooling rates could be obtained in the solid state than were otherwise available. The Duwez "gun" was soon followed by other similar devices in which the liquid was squeezed between a hammer and an anvil or between two oppositely moving pistons, and for some time the method became known as "splat cooling". However, only milligram samples could be treated in this way. Calculations and experiments (Duflos and Cantor, 1982) have shown that the effective rate of cooling of iron specimens in the piston and anvil techniques ranged from  $10^6$ - $10^7$  K s<sup>-1</sup> at 1500 K to  $10^1$ - $10^3$  K s<sup>-1</sup> at 500 K. The measured variation with temperature showed that it is not possible to quote a single cooling rate for a particular technique.

The difficulty of the small quantities of material was solved a little later when the technique of melt spinning was first used. The liquid is sprayed on to the surface of a rotating copper drum at a particular point on the circumference and is removed as a continuous solid layer at another fixed point on the circumference. This method has the enormous advantage of producing a continuous strip of treated material of appreciable width and indefinite length, but the cooling rate is inevitably lower; according to some measurements made by Hayzelden (1983) for a 40 nm steel ribbon, the cooling rate is  $3 \times 10^5$  K s<sup>-1</sup> at 1500°C but only 104 K s<sup>-1</sup> at 1000°C because the ribbon has lost its intimate contact with the drum. These figures show that the cooling rate varies markedly with temperature and emphasize that an estimate of the overall rate, much quoted in the early days of the technique, is without significance; what matters is the rate of cooling through some critical temperature range. Various other techniques, e.g. a pendant drop

method and a double drum spinner, have been reported, but appear to have no significant advantages over the two piston technique and the single drum melt spinner.

In some applications of rapid solidification, only the surface layers of a component have to be treated. This is accomplished by melting a thin layer with a continuous acting laser which is scanned over the surface or with a pulsed laser. Rapid cooling of the thin melted layer then results because of its intimate contact with the bulk component. A survey of the various techniques for rapid solidification is given in a review by Suryanarayana (1991) and in the book by Anantharaman and Suryanarayana (1987).

Rapid solidification produces specimens with properties which are in many ways quite different from normally treated materials. The changes are largely the result of lack of sufficient time for diffusion to occur at any temperature during the cooling. Thus microsegregation is almost eliminated and the specimens are very homogeneous. Another factor is the very small grain sizes – down to a few nanometers – found in rapidly solidified specimens, indicating copious nucleation and relatively slow growth. Solubility limits are greatly extended and two-phase fields correspondingly contracted. In the limit, as found by Duwez for copper–silver alloys, a continuous solid solution may be formed in a system which at equilibrium consists of a eutectic mixture with strictly limited solubility of each metal in the other. The small grain size of the rapidly solidified austenite was suggested by Inokuti and Cantor (1979) to be, in part at least, responsible for the depression of the martensitic  $M_s$  temperature in iron–nickel alloys, which they found increases with solute content and reaches 200 K in an iron–18% nickel alloy. In later experiments, Hayzelden and Cantor (1982) found that the structure was martensitic after quenching conventionally or rapidly using the two piston technique but was massive after melt spinning. This is an apparently anomalous result because the cooling rate of melt spinning is intermediate between those of the other two methods. It was explained (see Fig. 24.1) by postulating rather large depressions of  $M_s$  by rapid cooling compared with a relatively small grain size effect and this allowed the cooling curve of the melt-spun specimens to cross the curves of the massive transformation at temperatures above the depressed  $M_s$  (see Fig. 24.1). No reason was advanced to explain the large dependence of  $M_s$  on the cooling rate: thermally activated nucleation is a possibility but would be expected equally to depress the massive transformation. The results of this investigation also supported the concept that the temperature  $T_0$  of equal free energies forms the limit to massive transformation. The martensitic transformation in zirconium and in zirconium–niobium alloys is suppressed in rapidly cooled specimens and is replaced by the  $\alpha \rightarrow \omega$  transformation (Bannerjee and Cantor, 1979).

Rapidly cooled alloys not infrequently solidify in an amorphous form and this is discussed in Section 100. Cahn and Greer (1991) consider the freezing of an alloy occurring without any changes in composition and they divide the temperature range into three regions. Region A is above the melting temperature where, in metals at least, the viscosity is low and the atoms are highly mobile. In region C below the glass transition temperature, the liquid has a very high viscosity and atomic mobility is virtually zero, so that no further structural changes can take place. Region B is where the action is. If the specimen can reach the glass transition temperature without nucleating crystalline solid or

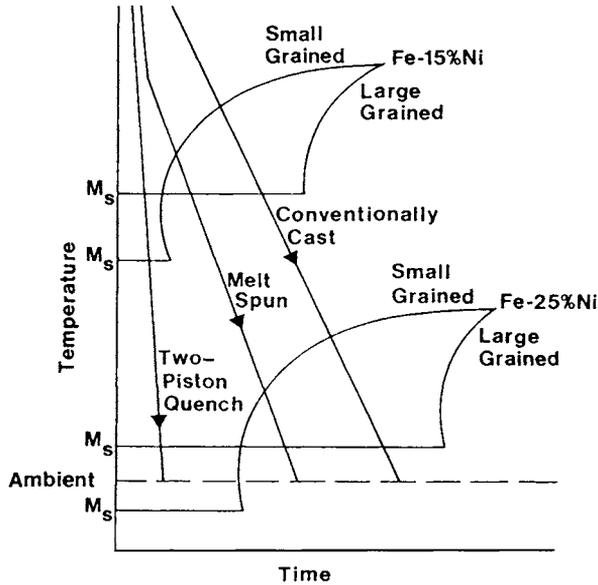


FIG. 24.1. Schematic  $T-T$  curves for massive transformation from austenite to ferrite and superimposed cooling curves for conventionally cast and rapidly solidified iron-nickel alloys (after Hayzelden and Cantor, 1982).

without nuclei being able to grow, then the solid formed will be amorphous. This requires a very high cooling rate over region B. The glass transition temperature is a kinetic parameter which varies, for example, slightly with cooling or heating rate.

### 100. AMORPHOUS (GLASSY) METALS

Metals were first made in the amorphous state by vapour deposition or chemical reaction or electrodeless deposition; the exact dates are not clear because the structures obtained were not classified as amorphous at the time. Duwez (1967) has given a brief history of these experiments which he traced back to a paper given by Wurtz in 1845, but the first clear recognition that a metal was being produced in an amorphous form came from the work of Buckel and Hilsch (1952), who worked with tin deposited from the vapour at very low (liquid helium) temperatures; this work was described by Buckel (1984) at the Fifth Conference on Rapidly Quenched Metals held in Würzburg in 1984.

Metallurgists were not familiar with this work and it came as a considerable surprise to many of them when the first experiments by Duwez showed that some alloys could be rapidly cooled into an amorphous or glassy state. The term glassy seems more appropriate as it emphasizes that the material has formed directly from the liquid and may not necessarily have the same structure as amorphous solids prepared in various other ways. Actually, the structures of amorphous solids are usually very similar to one another and to

the liquid phase, but there are exceptions. Cahn and Greer (1991) cite the case of silicon, which has metallic bonding in the liquid state but covalent bonding in the crystalline solid state and in the amorphous solid state.

Many methods of making amorphous solids are now available, including the application of high pressures, simple annealing below the melting point, interdiffusion of the constituents and ball milling of the constituent powders.

No pure metal has ever been obtained in the amorphous state by rapid cooling from the liquid. There are several groups of alloys which form glasses readily in rapidly cooled samples and their compositions are usually close to a deep eutectic in the equilibrium diagram, although the range of compositions is asymmetric with respect to the composition of the eutectic itself. It is normally easier to promote glassy behaviour by using ternary or multicomponent alloys; however, among binary alloys, there are several well-recognized regions in which it is possible to obtain specimens in the amorphous state. The first of these is a combination of a late transition metal with a metalloid in the range 15–25% metalloid. Another group consists of two transition metals, one from early in the series and one late. These particular alloys are the main alloys to be made amorphous by solid-state reaction. Other typical binary alloys are listed in Table XXV, which is taken from a review by Cahn and Greer (1991). Various models have been proposed to predict the ranges of composition of alloys which form glassy phases; the simplest and possibly most successful criterion is based on the size difference of the atoms in an alloy; this must be greater than about 15% for successful formation of amorphous alloys.

One measure of the structure of a liquid or a glass is the so-called free volume which is defined as the unfilled space between the atoms. The main concern is the volume additional to that which would be expected in ideal close packing of spheres. It is assumed that the smaller the free volume, the less mobile are the atoms. The viscosity varies with the relative free volume  $f$  defined as the change in volume between 0 K and the temperature considered divided by the volume at 0 K. Some glasses obey an equation for the viscosity  $\eta$

$$\eta = A \exp(B/f)$$

which applies throughout region  $B$ .

Amorphous alloys formed in the solid state or by vapour condensation appear to be structurally very similar to glasses. A dramatic example of the continuity of the liquid and amorphous alloy produced in the solid state is provided by chromium titanium alloys. As prepared normally, there is a continuous b.c.c. solution at temperatures just below the solidus. However, when an alloy of 45 at.% Ti is annealed, the structure is converted to an amorphous structure of the same composition. The question then arises is this amorphous phase to be regarded thermodynamically as the same phase as the liquid? If so, there must be two intersections of the two free energy curves of the b.c.c. phase and the liquid phase curve must be something like Fig. 24.2. These curves were actually calculated using the CALPHAD scheme and the corresponding phase diagram is shown in Fig. 24.3. There are only three phases in Fig. 24.3, the liquid and amorphous regions being regarded as essentially the same phase. Notice that at  $T_{im}$  on Fig. 24.2 there is an “inverse melting point” (this term is due to Greer, 1988), i.e. the amorphous phase should form on cooling

TABLE XXV. EXAMPLES OF GLASS-FORMING ALLOYS,<sup>1</sup> CLASSIFIED INTO THE CATEGORIES DISCUSSED IN THE TEXT. (after CAHN AND GREER, 1996)

Type	Alloy	Glass-forming range	References
(i) Late transition metal-metalloid	$\text{Fe}_{100-x}\text{B}_x$	$x = 12-25$	HASEGAWA and RAY [1978]
	$\text{Pd}_{100-x}\text{Si}_x$	$x = 14-22$	LEWIS and DAVIES [1976]
	$\text{Ni}_{100-x}\text{B}_x$	$x = 17, 18.5, 31-41$	Donald and DAVIES [1980]
	$\text{Pt}_{100-x}\text{Sb}_x$	$x = 34, 36.5$	SRIVASTAVA <i>et al.</i> [1972]
	$\text{Fe}_{80}\text{P}_{13}\text{C}_7$		DUWIZ [1967]
	$\text{Fe}_{40}\text{Ni}_{40}\text{B}_{20}$		LUBORSKY [1977]
	$\text{Fe}_{40}\text{Ni}_{40}\text{P}_{14}\text{B}_6$		POLK and CHEN [1974]
	$\text{Co}_{70.5}\text{Fe}_{4.5}\text{Si}_{15}\text{B}_{10}$		MAKINO <i>et al.</i> [1990]
	$\text{Pd}_{40}\text{Ni}_{40}\text{P}_{20}$		MATREPIERRE [1970]
(ii) Early transition metal-metalloid	$\text{Ti}_{100-x}\text{Si}_x$	$x = 15-20$	INOUE and MASUMOTO [1980]
	$\text{Re}_{65}\text{Si}_{35}$		FISCHER <i>et al.</i> [1978]
	$\text{W}_{80}\text{N}_{20}$		DÖRNER <i>et al.</i> [1991]
	$\text{W}_{60}\text{Ir}_{20}\text{B}_{20}$		FISCHER <i>et al.</i> [1978]
(iii) Early transition metal-late transition metal	$\text{Nb}_{100-x}\text{Ni}_x$	$x = 40, 70$	RUHL <i>et al.</i> [1967]
	$\text{Cu}_{100-x}\text{Zr}_x$	$x = 25, 60$	RAY <i>et al.</i> [1968]
	$\text{Ni}_{100-x}\text{Zr}_x$	$x = 10, 12, 33, 80$	CLEMENS <i>et al.</i> [1984], GREGAN <i>et al.</i> [1981]
	$\text{Fe}_{100-x}\text{Zr}_x$	$x = 9, 72, 76$	Horváth <i>et al.</i> [1988]
	$\text{Ta}_{100-x}\text{Ni}_x$	$x = 40, 70$	GIESSEN <i>et al.</i> [1976]
	$\text{Ta}_{55}\text{Rh}_{45}$		FISCHER <i>et al.</i> [1978]
	$\text{Zr}_{60}\text{Ni}_{25}\text{Al}_{15}$		INOUE <i>et al.</i> [1990]
(iv) Aluminum-based	$\text{Al}_{100-x}\text{La}_x$	$x = 10, 50, 80$	INOUE <i>et al.</i> [1988b]
	$\text{Al}_{70}\text{Ni}_{15}\text{Si}_{15}$		DUBOIS <i>et al.</i> [1992]
	$\text{Al}_{75}\text{Cu}_{15}\text{V}_{10}$		Holzer and KELTON [1991]
	$\text{Al}_{85}\text{Ni}_{10}\text{Zr}_5$		DUBOIS <i>et al.</i> [1992]
	$\text{Al}_{80}\text{Ni}_{10}\text{Y}_{10}$		INOUE <i>et al.</i> [1988a]
	$\text{Al}_{85}\text{Fe}_{10}\text{La}_5$		INOUE <i>et al.</i> [1988a]
	$\text{Al}_{85}\text{Ni}_5\text{Y}_8\text{Co}_2$		INOUE <i>et al.</i> [1990]
(v) Lanthanide-based	$\text{La}_{100-x}\text{Al}_x$	$x = 18, 34$	AGYEMAN <i>et al.</i> [1979]
	$\text{La}_{100-x}\text{Ge}_x$	$x = 17, 22$	AGYEMAN <i>et al.</i> [1979]
	$\text{La}_{100-x}\text{Au}_x$	$x = 18, 26$	JOHNSON <i>et al.</i> [1975]
	$\text{Gd}_{100-x}\text{Fe}_x$	$x = 32, 50$	BUSCHOW [1979]
	$\text{La}_{55}\text{Al}_{25}\text{Ni}_{20}$		INOUE <i>et al.</i> [1989]
(vi) Alkaline-earth	$\text{Mg}_{100-x}\text{Zn}_x$	$x = 25, 32$	CAIKA <i>et al.</i> [1977]
	$\text{Ca}_{100-x}\text{Al}_x$	$x = 12.5, 47.5$	SOMMER <i>et al.</i> [1978]
	$\text{Mg}_{65}\text{Cu}_{25}\text{Y}_{10}$		KIM <i>et al.</i> [1990]
	$\text{Be}_{40}\text{Zr}_{10}\text{Ti}_{50}$		HASEGAWA and TANNER [1977]
(vii) Actinide-based	$\text{U}_{100-x}\text{Co}_x$	$x = 24, 40$	RAY and MUSSO [1976]

<sup>1</sup>All compositions are quoted in at.%. The amorphous alloys have been formed by a variety of techniques, mostly rapid liquid quenching. The compositions given are those of representative alloys and do not imply that amorphous phase formation at other compositions in the same system would not be possible. For binary alloys, however, some indication is given of glass-forming ranges.

References: pp. 1087 + 1088

because it is more stable than the crystalline solid solution. The actual structure of the amorphous phase is as difficult to model as is the structure of liquids (see Section 21). X-ray and neutron techniques usually give only two diffuse diffraction peaks, from which the radial distribution function may be determined by Fourier inversion of the scattering

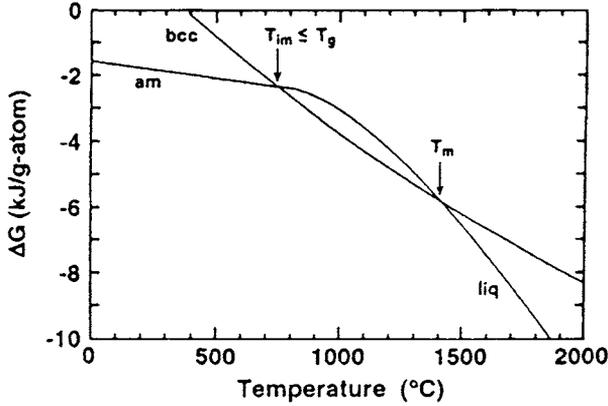


FIG. 24.2. Calculated free energy versus temperature curves for the liquid/amorphous phase and the crystalline solid solution in a chromium 45 at.% titanium alloy (after Bormann, 1994). The ordinary melting temperature is  $T_m$  and the “inverse melting temperature” is  $T_{im}$ .

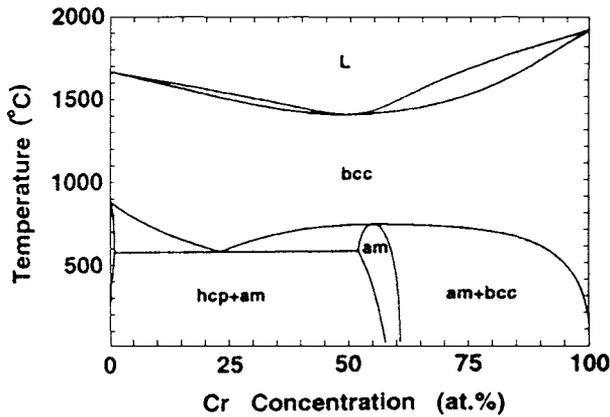


FIG. 24.3. Calculated metastable phase diagram for the chromium-titanium binary system when only the b.c.c. and h.c.p. solid solutions and the amorphous phase are allowed (after Bormann, 1994). The liquid and the solid amorphous phase are considered to be thermodynamically the same phase.

data and the mean coordination number may be deduced from the area under the curve. By using radiations of different atomic scattering powers, partial radial distribution functions may be obtained. In this way, it has been shown that metalloid atoms, when present, tend to avoid each other so that metalloid contacts are rare. The average number of near neighbours around a central metalloid atom is usually eight to nine and around a central metallic atom it is 10–12. Theories of these structures as crystal lattice plus a high concentration of vacancies or a very small grain size do not work; the individual grains of

a polycrystalline assembly for example retain their identity right into the nanocrystalline region. X-ray diffraction patterns may now be obtained much more rapidly and probably more accurately by use of synchrotron radiation.

Models of amorphous and liquid structures were initially based on Bernal's concept of dense random packing of spherical atoms. In linking this to free volume concepts, he considered (see p. 162) the polygonal shapes which represent holes which could not accommodate spherical atoms. He proposed that the structure contains five such polyhedra and for a time this was the most favoured structural model of glass. However, an alternative model due to Gaskell (1983) considers that the atoms surround trigonal prisms which are joined together in one of two ways. This model was further elaborated by Dubois *et al.* (1985), who introduced the idea of chemical twinning on a very fine scale (1–2 nm). Some rather similar models have come from experiments on pressure-induced amorphization. They are based on the concept of tetrahedral packing of four hard spheres. This is clearly a minimum energy configuration for a cluster containing only four atoms.

Tetrahedra cannot fill space and although other polyhedra may be formed from slightly distorted tetrahedra (e.g. icosahedra) these often have fivefold symmetries and so cannot be part of a space lattice, so that in any structure built from slightly distorted tetrahedra their fivefold symmetry prevents the establishment of translational symmetry. The amorphization of minerals under pressure is discussed by Sharma and Sikka (1996); some semi-metals amorphize under decompression rather than under static or dynamic compression.

Structural models of these general types seem at present to be the best descriptions available for liquids and glassy metals although they still do not give radial distribution curves in exact agreement with experiment. Various other techniques have been used to help in modelling the structure; like the diffraction studies, all of these methods only give one-dimensional information. This difficulty may be overcome by the use of field ion microscopy with a position-sensitive atom probe. This, in principle, allows the position of each atom and its chemical species to be uniquely defined. In particular, the angular distribution of nearest-neighbour contact may be deduced and this information is not obtainable in any other way. De Hossien (1983) measured the planar radial distribution functions from field ion images of the palladium atoms in an Fe-Cu-Si-Pd glass, and also the angular dependence of nearest-neighbour interatomic vectors. The latter distribution has maxima at 60°, 90°, 120° and 150°. It is suggested that the strongest peak at 60° is evidence of tetrahedral packing, but the other peaks suggest that other polyhedra may be involved. Field ion microscope studies of amorphous alloys have been also made by Hono and co-workers (1991, 1992) and by Oehring and Haasen (1986). In some alloys, clustering was observed within the amorphous solid solution prior to crystallization; in others no effects of this kind were seen.

Changes in physical properties occur when a metallic glass is annealed in the region of its glass transition temperature. These changes, which are called relaxation, occur in order to change the structure of the glass to the metastable equilibrium structure which it was unable to obtain during the rapid cooling. One such change is produced by the reduction in the vacancy concentration towards its equilibrium value. This in turn produces a reduction in the diffusion coefficient and the atomic volume or free volume. Some changes

like the one just cited are irreversible in the sense that the pre-annealing value of the property can never be obtained again, but others (for example, elastic moduli) can behave reversibly on changing the temperature after the original change has saturated. These are referred to as monotonic and reversible relaxations respectively. Egami (1978) suggested that monotonic relaxation involves topological short-range ordering in which there is a rearrangement of atomic sites and a reduction in the free volume whilst reversible relaxation involves composition (chemical) short-range ordering. This latter process is supposed to be a rearrangement of atoms of different species on a fixed array of sites and can occur at lower temperatures than topological ordering.

Crystallization of a glass ("devitrification") can be studied in many ways more readily than solidification because of the high driving force and low atomic mobility. The temperature range within which crystallization occurs in reasonable periods is typically about 45% of the melting temperature.

Both the rate of nucleation of a crystalline phase from a supercooled liquid and the subsequent growth rate of these nuclei exhibit a maximum as the temperature is lowered below the freezing point. This maximum arises because of the increasing driving force and the decreasing atomic mobility as the temperature is reduced. Solidification normally takes place in a temperature range immediately below the solidus; in rapidly solidified materials the formation of a crystalline phase is either prevented completely or is depressed to much lower temperatures in a range just above the glass transition temperature. Although the two reactions are in principle the same, crystallization of a glass is very much easier to control as it can readily be halted at any stage by sudden cooling. As in the freezing of a liquid, there are various possibilities for the way in which an amorphous solid devitrifies, some of which are illustrated in Fig. 24.4. The reactions possible include polymorphic crystallization, in which a crystalline phase of the same composition as the glass is formed, primary crystallization, in which a crystalline phase of a different composition is formed and eutectic crystallization, in which crystalline phases of different compositions grow cooperatively from the amorphous phase. Measurements of nucleation and growth rates may be made more readily during the crystallization of a glass than in the solidification of a liquid simply because the reaction may be terminated at any stage by rapid cooling to below the glass transition temperature.

Nucleation without growth may occur near to the glass transition temperature during the rapid cooling to obtain the initial amorphous structure. On annealing at temperatures where crystallization takes place, these nuclei may grow and, if no others form, the number of nuclei in any given volume will be constant with time. Alternatively, nuclei may form during the annealing treatment, either homogeneously or heterogeneously. (Evidence has been obtained for both.) The nucleation rate may be steady state or there may be quite a lengthy transient. Some schematic nucleation curves of a number of product regions versus time are shown in Fig. 24.5.

Nucleation rates in an  $\text{Fe}_{40}\text{Ni}_{40}\text{P}_{14}\text{B}_6$  glass have been analysed by Kelton (1991). The experimental data are roughly of type (b) in Fig. 24.5, indicating homogeneous nucleation, but there is considerable difficulty in fitting the results numerically. Standard nucleation theory [see eqn. (49.14)] leads to the result that the logarithm of the product

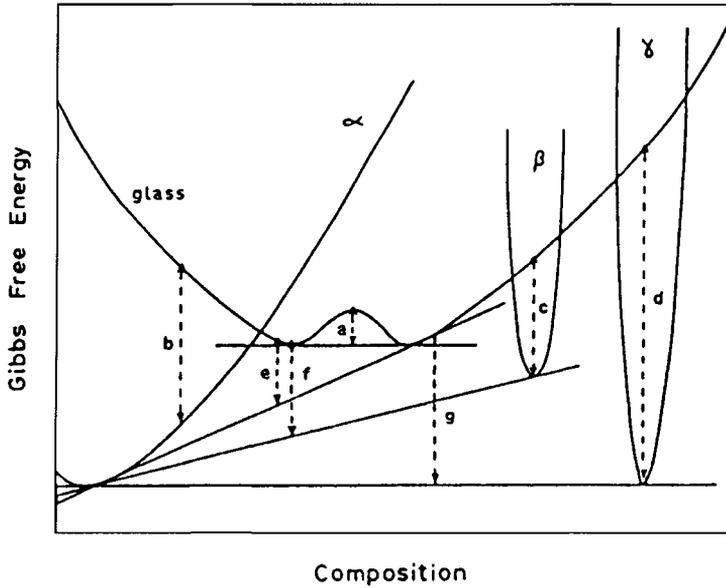


FIG. 24.4. Some possible crystallization reactions in a binary system. Free energy curves to show driving force for amorphous phase: (a) to separate into two amorphous phases of different compositions; (b) to transform polymorphically into a supersaturated solid solution; (c) to transform into a metastable phase  $\beta$ ; (d) to undergo a similar change into a stable phase  $\gamma$ ; (e) to undergo primary crystallization of  $\alpha$ ; (f) to crystallize as a metastable eutectic  $\alpha + \beta$ ; and (g) to form an equilibrium eutectic  $\alpha + \gamma$ .

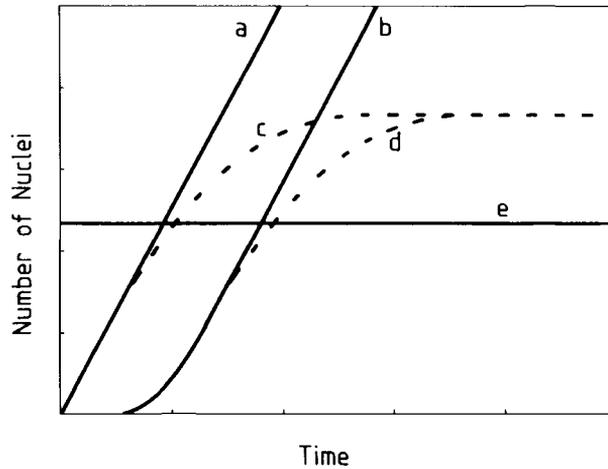


FIG. 24.5. Schematic variation with time of the number of product regions per unit volume produced by (a) steady-state homogeneous nucleation; (b) steady-state heterogeneous nucleation; (c, d) corresponding transient nucleation; and (e) all nuclei present initially. (After Cahn and Greer, 1991.)

of the nucleation rate per unit volume and the viscosity should vary linearly with the supercooling  $\Delta T^-$ , provided that the surface free energy is constant and isotropic. The experimental data support this interpretation over much of the range of  $\Delta T^-$  but there are deviations at high and low values of the supercooling. The pre-exponential derived from the data is reasonable for homogeneous nucleation but, if the effective surface free energy is allowed to vary in such a way as to satisfy the experimental observations, the pre-exponential constant has a much smaller value, more typical of heterogeneous nucleation.

Studies of the crystallization of amorphous materials by field ion microscopy have been made by Hono and his co-workers (1991, 1992) and by Piller and Haasen (1982). Hono and co-workers observed concentration fluctuations at an early stage in the process, and this is believed to lead to a large density of nuclei. During subsequent annealing, they suggest that the crystal growth is controlled by one or two of the following mechanisms: (1) stabilization of the amorphous phase by partitioning of amorphous forming elements; (2) segregation of a slow diffusing solute at the interface between the amorphous phase and the crystalline phase; and (3) precipitation of a thermally stable compound.

### 101. QUASI-CRYSTALS

A brief description of quasi-crystals was given in Chapter 13; to supplement this, it is emphasized here that quasi-crystals are highly ordered structures, even though they lack translational periodicity. Quasi-crystals may be regarded as a type of intermetallic compound in which quasi-periodic translational order (see p. 617) is imposed. Their structures and properties have been reviewed by Kelton (1993) and by Ranganathan and Chattopadhyay (1991). The lack of translational periodicity does not prevent the development of either (chemical) order or twinning, although naturally those properties which depend on long-range periodicity are not exhibited.

Mukhopadhyay *et al.* (1987, 1989) first showed that the crystals could order to form superlattice structures and Henley (1988) showed that the icosahedral ordering could be related to the formation of a f.c.c. hypercube in six-dimensional space, whereas an unordered structure is obtained by projection of a simple hypercube in six-dimensional space. A stable icosahedral structure with the appropriate ordering was found soon afterwards in an Al-Cu-Fe alloy (Tsai *et al.*, 1987).

Twinning of the icosahedral structure was reported by Koskenmaki *et al.* (1986). Rosette-shaped icosahedral grains were twinned with a mirror plane normal to one of the fivefold axes. The combined symmetry of such a twinned grain is identical with that of a decagonal quasi-crystal.

Much work in the 1990s concerns the decagonal structure and this has been reviewed by Ranganathan *et al.* (1997). The decagonal phase has point group symmetry  $10mmm$  and has a unique tenfold symmetry axis. It has been found only in aluminium alloys with the solitary exception of an alloy of zinc with magnesium and a rare earth metal. Investigations of its structure include data from X-ray and electron diffraction and from

high resolution electron microscopy. An electron micrograph taken down the tenfold axis shows a pattern of pentagonal rings (Hiraga *et al.*, 1978).

Models of the structure are of two kinds. One set is based on atomic clusters with fivefold symmetry; such models are often developed from rather similar crystalline compounds. An alternative approach is to use decorated Penrose tiles in layers perpendicular to the tenfold axis and stacked in various sequences. This model has been compared to polytypism in crystalline materials. The repeat distances along the tenfold axis are variously about 4, 6, 8, 12 and 16 Å in different compounds and in some cases in the same compound. Figure 24.6 shows a model structure for  $\text{Al}_3\text{Fe}$  with two layers of

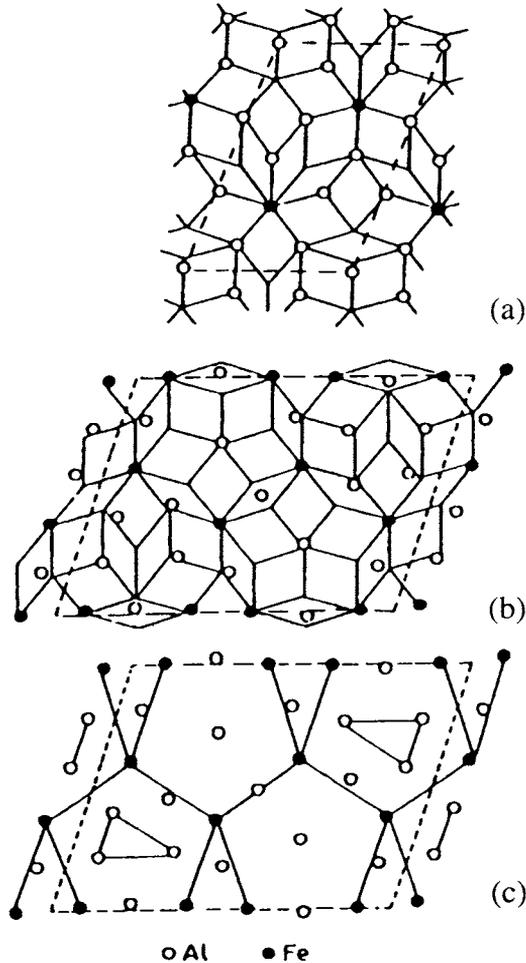


FIG. 24.6. Model of the  $\text{Al}_3\text{Fe}$  decagonal quasi-crystal as two layers, (a) and (b), each consisting of decorated Penrose tiles normal to the tenfold axis (after Kumar *et al.*, 1986). Part (c) shows how the wide and narrow rhombic tiles in layer (b) can be redrawn as pentagons and narrow rhombic tiles; this is similar to the pentagonal rings observed in high resolution electron micrographs down the tenfold axis.

tiling by wide and narrow rhombus tiles. Part (c) demonstrates how the layer (b) may alternatively be regarded as tiling of pentagonal and narrow rhombus tiles and this is very similar to patterns observed by high resolution electron microscopy.

## REFERENCES

- AGYEMAN, K., MULLER, R., and TSUEI, C. C. (1979) *Phys. Rev. B* **19**, 193.
- ANANTHARAMAN, T. R. and SURYANARAYANA, C. (1987) *Rapidly Solidified Metals*, Trans. Tech. Publications, Zürich, Switzerland.
- BANNERJEE, S. and CANTOR, B. (1979) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-79)*, p. 195, MIT Press, Cambridge, MA.
- BORMANN, R. (1994) *Mater. Sci. Eng. A* **179-180**, 31.
- BUCKEL, W. (1984) *Rapidly Solidified Metals*, Vol. II, eds. S. Steed and H. Warlimont, p. XIX, North-Holland, Amsterdam.
- BUCKEL, W. and HILSCH, R. (1952) *Z. Phys.* **132**, 420.
- CAHN, R. W. and GREER, A. L. (1991) *Physical Metallurgy*, 4th edn., eds. R. W. Cahn and P. Hassen, p. 172, North-Holland, Amsterdam.
- CAJKA, A., MADHAVA, M., POLK, D. E., GIESSEN, B. C., MATYJA, H., and VAN DER SANDE, J. B. (1977) *Scr. Metall.* **11**, 65.
- DE HOSSEN, J. Th. M. (1983) *Amorphous Materials: Modeling of Structure and Properties*, ed. V. Vitek, p. 147, Metallurgical Society of AIME, Warrendale, PA.
- DONALD, I. W. and DAVIES, A. H. (1980) *J. Mater. Sci.* **15**, 2754.
- DÖRNER, W., MEHRER, H., PKELA, P. J., KOWALA, E., and NICOLET, M.-A. (1991) *Mater. Sci. Eng. B* **10**, 165.
- DUBOIS, J. M., GASKELL, P. H., and LE-CAËR, G. (1985) *Rapidly Quenched Metals V*, eds. S. Steeb and H. Warlimont, p. 567, North-Holland, Amsterdam.
- DUFLOS, B. and CANTOR, B. (1982) *Acta Metall.* **30**, 323.
- DUWEZ, P. (1967) *Trans. Am. Soc. Met.* **60**, 607.
- DUWEZ, P., WILLIAMS, R. R., and KLEMENT, M. (1960) *J. Appl. Phys.* **31**, 1136.
- EGAMI, T. (1978) *Mater. Res. Bull.* **13**, 557.
- FISCHER, M., POLK, D. E., and GIESSEN, B. C. (1978) *Proc. 1st Int. Conf. Rapid Solidification Processing*, eds. R. Mehrabian, B. H. Kear and M. Cohen, p. 140, Claitors, Baton Rouge, LA.
- GASKELL, P. H. (1983) *Amorphous Materials: Modeling of Structure and Properties*, ed. V. Vitek, p. 47, Metallurgical Society of AIME, Warrendale, PA.
- GIESSEN, B. C., MADHAVA, M., POLK, D. E., and VAN DER SANDE, J. B. (1976) *Mater. Sci. Eng.* **23**, 145.
- GREER, A. L. (1988) *J. Less-Common Met.* **140**, 327.
- HASEGAWA, D. R. and TANNER, L. E. (1977) *J. Appl. Phys.* **49**, 3211.
- HASEGAWA, S. and RAY, R. (1978) *J. Appl. Phys.* **49**, 4174.
- HAYZELDEN, C. (1983) *D. Phil. Thesis*, Sussex University, Brighton, UK.
- HAYZELDEN, C. and CANTOR, B. (1982) *Proc. Int. Conf. Solid-Solid Phase Transformations, Pittsburgh, PA, 1981*, eds. H. I. Aaronson, D. E. Laughlin, R. F. Sekerka and C. M. Wayman, p. 1397, The Metallurgical Society of AIME, Warrendale, PA.
- HENLEY, C. L. (1988) *Philos. Mag. Lett.* **58**, 87.
- HIRAGA, K., LINCOLN, F. J., and SUN, W. (1978) *Mater. Trans. Jpn. Inst. Met.* **32**, 308.
- HONO, K., HIRAGA, K., WANG, Q., INOUE, A., and SAKURAI, T. (1992) *Acta Metall.* **40**, 2137.
- HONO, K., INOUE, A., and SAKURAI, T. (1991) *Appl. Phys. Lett.* **58**, 2180.
- HORVATH, J., OTT, J., PEABLER, K., and ULFERT, W. (1988) *Mater. Sci. Eng.* **97**, 409.
- INOKUBI, Y. and CANTOR, B. (1979) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-79)*, p. 76, MIT Press, Cambridge, MA.
- INOUE, A., MASUMOTO, N., and MASUMOTO, T. (1990) *Mater. Trans. Jpn. Inst. Met.* **31**, 493.
- INOUE, A. and MASUMOTO, T. (1980) *Sci. Rep. Res. Inst. Tohoku Univ., Ser. A* **28**, 165.
- INOUE, A., ZHANG, T., and MASUMOTO, T. (1989) *Mater. Trans. Jpn. Inst. Met.* **30**, 965.

- JOHNSON, W. L., POON, S. J., and DUWEZ, P. (1975) *Phys. Rev. B* **11**, 150.
- KELTON, K. F. (1991) *Solid State Phys.* **85**, 55.
- KIM, Y. H., INOUE, A., and MASUMOTO, T. (1990) *Mater. Trans. Jpn. Inst. Met.* **31**, 747.
- KLEMENT, M., WILLIAMS, R. R., and DUWEZ, P. (1960) *Nature(London)* **187**, 869.
- KOSKENMAKI, D. C., CHEN, H. S., and RAO, K. V. (1986) *Scr. Metall.* **20**, 1631.
- KUMAR, V., SCHIOO, D., and ATHITHANC, G. (1986) *Phys. Rev.* **34**, 6924.
- LEWIS, B. G. and DAVIES, H. A. (1976) *Mater. Sci. Eng.* **23**, 179.
- LUBORSKY, F. E. (1977) *Amorphous Metallic Alloys*, ed. F. E. Luborsky, p. 360, Butterworths, London.
- MAITREPIERRE, P. L. (1970) *J. Appl. Phys.* **41**, 498.
- MAKINO, A., INOUE, A., and MASUMOTO, T. (1990) *Mater. Trans. Jpn. Inst. Met.* **31**, 884.
- MUKHOPADHYAY, N. K., RANGANATHAN, S., and CHATTOPADHYAY, K. (1987) *Philos. Mag. Lett.* **56**, 121; (1989) *Ibid.* **60**, 207.
- OEHRING, M. and HAASEN, P. (1986) *J. Phys. (Paris), Colloq.* **47-C7**, 275.
- PILLER, J. and HAASEN, P. (1982) *Acta Metall.* **30**, 1.
- POLK, D. E. and CHEN, H. S. (1974) *J. Non-Cryst. Solids* **15**, 165.
- RANGANATHAN, S. and CHATTOPADHYAY, K. (1991) *Ann. Rev. Mater. Sci.* **21**, 437.
- RANGANATHAN, S., CHATTOPADHYAY, K., SINGH, A., and KELTON, K. F. (1997) *Prog. Mater. Sci.* **41**, 195.
- RAY, R. and MUSSO, E. (1976) *US Pat.* 3 981 722.
- RUHL, R. C., GIessen, B. C., COHEN, M., and GRANT, N. J. (1967) *Acta Metall.* **15**, 1693.
- SHARMA, S. M. and SIKKA, S. K. (1996) *Prog. Mater. Sci.* **40**, 1.
- SOMMER, F., DUDDEK, G., and PREDEL, B. (1978) *Z. Metallkd.* **69**, 587.
- SURYANARAYANA, C. (1991) *Processing of Metals and Alloys (Materials Science and Technology, Vol. 15)*, eds. R. W. Cahn, P. Haasen and E. J. Kramer, p. 57, VCH, Weinheim.
- TSAL, A. P., INOUE, A., and MASUMOTO, T. (1987) *J. Mater. Sci Lett.* **6**, 1403.

## CHAPTER 25

# *Bainite in Steels*

### 102. MICROSTRUCTURES AND PROPERTIES

Bainite is the name given to the microstructural constituent formed in steels by the decomposition of austenite at temperatures above those of martensite formation but below those at which fine pearlite is formed. This transformation is not easy to understand and has given rise to much controversy, only an outline account of which can be given here. A fuller description is given in the book by Bhadeshia (2001) and in various review articles (Bhadeshia and Christian, 1990; Christian and Edmonds, 1984; Hehemann, 1970; Aaronson, 1962, 1969; Aaronson *et al.*, 1990).

Bainite formed above about 350 C (upper bainite) consists of ferrite plates or laths (subunits) and a dispersion of carbide precipitates. The ferrite plates are in aggregates called sheaves (Aaronson and Wells, 1956) and are separated from each other by regions of retained austenite or, after cooling, by transformed austenite, i.e. martensite. The carbide (usually cementite) is precipitated between the ferrite plates. In lower bainite, the carbide is precipitated inside the ferrite; it takes the form of plate-shaped particles orientated at about 60° to the axis of the ferrite plates. Figures 25.1 and 25.2 are micrographs showing upper and lower bainite respectively.

The transition from upper to lower bainite occurs over a fairly narrow temperature range and apparently is not sensitive to composition (see Fig. 25.3, which shows the transition temperature as a function of carbon content). Electron microscope studies by Irvine and Pickering (1963) showed that the edges of ferrite plates in lower bainite have sharp radii of curvature and appear to be free of carbides, but the sides of a plate are nearly planar and carbides are then in contact with the interface. According to Matas and Hehemann (1961), the carbide first formed in hypoeutectoid steels is  $\epsilon$ -carbide, but this is later replaced by cementite. Thus lower bainite has undergone a decomposition sequence rather similar to that of tempered martensite. In hypereutectoid steels, carbide forms before ferrite, a situation which Hillert (1962) described as "inverse bainite". The initial cementite crystals appear as long laths or plates and, according to Pickering (1967), rapid side-by-side nucleation then leads to a structure in which laths of ferrite and cementite appear to grow alternately. However, in a detailed investigation, Kinsman and Aaronson (1970) showed that although bainite in hypereutectoid steels begins with long single spines of cementite, the region around any such spine develops into a structure resembling

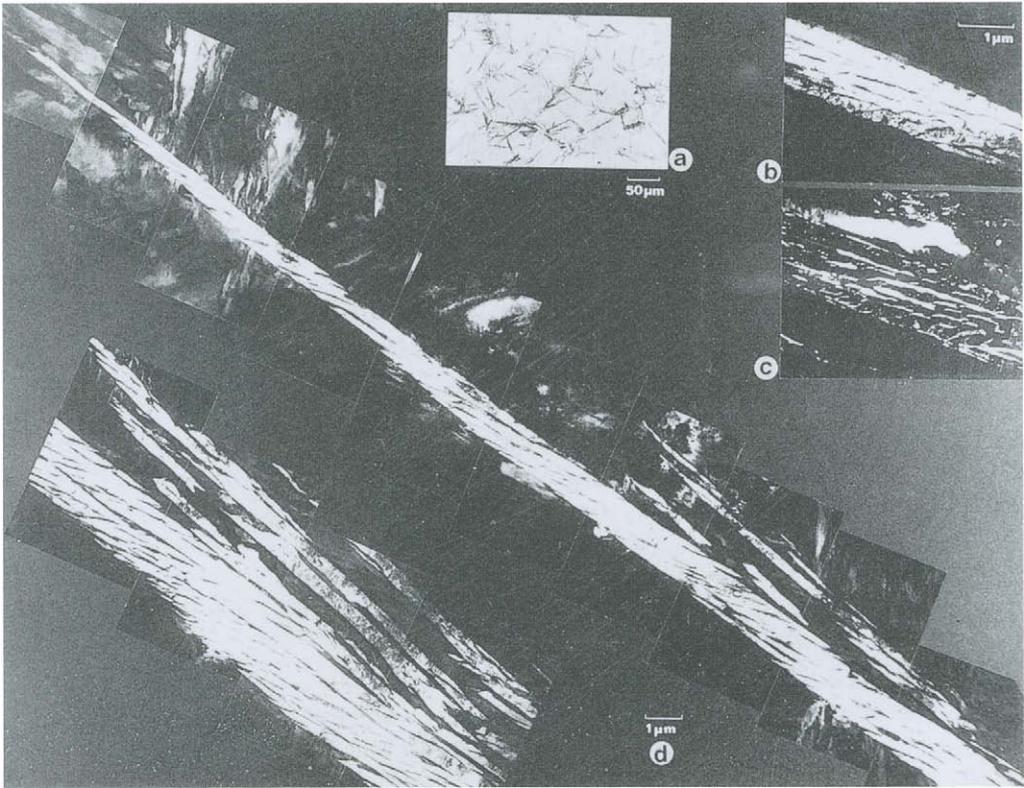


FIG. 25.1. (a) Microstructure of upper bainite in silicon steel isothermally transformed at 286 C for 30 min (after Bhadeshia and Edmonds, 1979, 1980). (b) Optical micrograph of sheaves of upper bainite in a martensitic matrix. (c) Corresponding retained austenite dark field image. (d) Montage of electron micrographs showing morphology and substructure of a sheaf of upper bainite.

normal bainite, with large ferrite laths and smaller carbide particles. Kinsman and Aaronson ascribed the limited ability of the initial inverse bainite structure to reproduce itself to the greater volume fraction and higher growth velocity of the ferrite regions.

A feature of the isothermal formation of bainite is that the transformation does not proceed to completion, but ceases after some volume fraction of the austenite has been transformed. This fraction increases with decreasing temperature and a plot of the fraction of austenite transformed versus the temperature may be extrapolated back to a temperature  $B_s$  at which the bainite reaction may be supposed to begin. The  $B_s$  temperature defined kinetically in this way generally agrees with a microstructural definition of the onset of bainite formation. The concept of a  $B_s$  temperature is strengthened by  $T-T-T$  curves for alloy steels which show a pronounced "bainitic bay" separating the pearlite and bainite reactions. The absence or extreme sluggishness of reaction in this temperature range is interpreted as its being above  $B_s$ , but an alternative

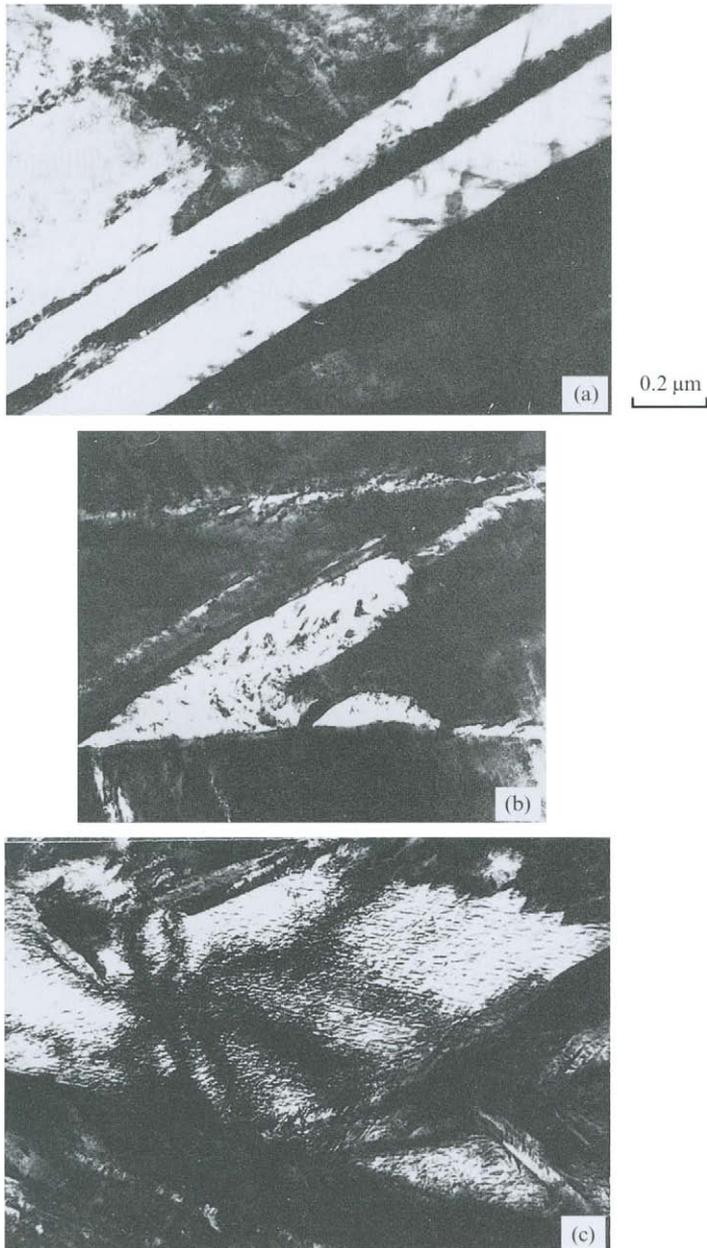


FIG. 25.2. (a) Spikes of lower bainite after isothermal transformation at 257 C for 5 min. (b) Spikes of grain boundary lower bainite after transformation at 237 C for 10 min. (c) Grain boundary lower bainite formed after isothermal treatment at 237 C for 10 min.

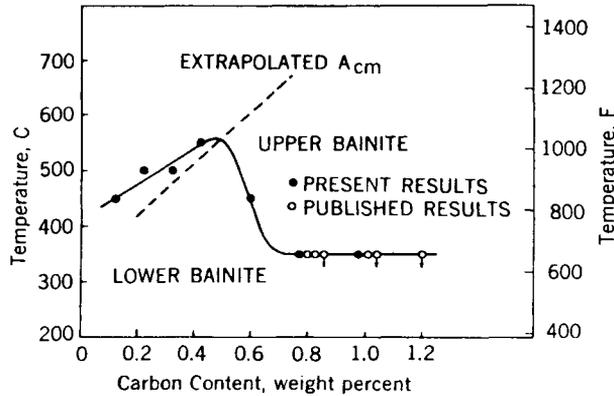


FIG. 25.3. Dependence of transition temperature between upper and lower bainite on carbon content (after Irvine and Pickering, 1963).

interpretation due to Aaronson (1969) attributes the effect to solute drag on the interface, and assigns no significance to the  $B_s$  temperature. Oblak and Hehemann (1967) emphasized that there is a marked difference between the structure of upper bainite, which consists of a parallel arrangement of ferrite plates or laths with a rather high dislocation density, and that of pro-eutectoidal ferrite, which has no lath or cellular substructure and a low dislocation density. In the case of silicon steels, no carbides are precipitated, but Bhadeshia and Edmonds (1980) showed that subunits separated by films of austenite can also be distinguished in these alloys, and the reaction is accepted by most workers as bainitic. Figure 25.1 shows the microstructural appearance of a sheaf of upper bainite in a silicon steel.

Some indirect evidence that the ferrite plates formed at 216 C may contain up to 0.3 wt.% carbon has been presented (Vasudevan *et al.*, 1958) but the work of Matas and Hehemann shows that the carbon content of the ferrite is usually less than that of  $\alpha$ -martensite formed by first-stage tempering of martensite (see p. 749). It seems probable that the freshly formed ferrite must contain appreciable amounts of carbon, perhaps almost as much as the original austenite. The supersaturation is removed by the carbide precipitation, but the low carbon content of the ferrite also suggests that there may be some partition of carbon between austenite and ferrite during the growth process.

At higher temperatures in the upper bainite range, cementite precipitates without prior formation of  $\epsilon$ -carbide. The cementite plates are larger and appear to form directly from the austenite; they are orientated parallel to the main growth direction rather than at 60° to it. Clearly there is marked carbon enrichment of austenite during the formation of upper bainite, which grows first as nearly carbon-free ferrite. At the highest temperatures in the upper bainite range, the ferrite plates merge to form large areas of ferrite surrounded by dense clusters of cementite precipitated from the supersaturated austenite. At somewhat lower temperatures, the carbide plates precipitate from the austenite alongside the ferrite.

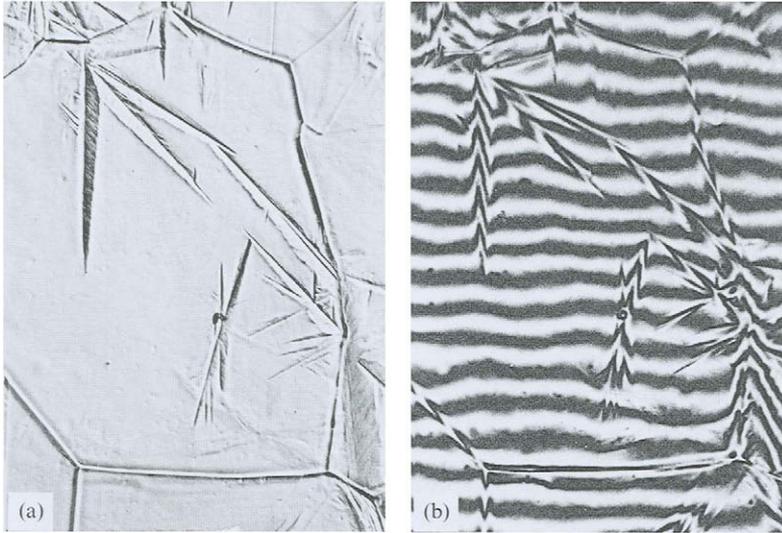


FIG. 25.4. Light micrograph (a) and interferogram (b) of relief effects accompanying the formation of bainite at 350 C (courtesy of Speich, 1962).

A significant observation is that bainite plates or laths are accompanied by an invariant plane strain shape change. The reaction is regarded as intermediate between nucleation and growth and martensitic types inasmuch as it exhibits slow isothermal growth and a shape change. The significance of the shape change was emphasized by Ko and Cottrell (1952) and these workers later suggested a growth mechanism which is related to martensite formation (Cottrell and Ko, 1953). Speich (1962) verified that the shape change is an invariant plane strain (see Fig. 25.4) although it must now be presumed that his observations related to a sheaf of ferrite rather than to individual subunits. Observations have recently been made of the shape change on a finer scale (Swallow and Bhadeshia, 1996) using atomic force microscopy and the tilt across an individual subunit can then clearly be distinguished (see Fig. 25.5). Swallow and Bhadeshia verified that the tilt across each subunit in a sheaf is identical, and they measured the maximum magnitude of the shape deformation to be 0.26, which is consistent with the theory of martensite crystallography (see Fig. 25.6).

The identical crystallography of the subunits in a sheaf raises the question are these distinct crystals or are they branches which are all linked together in much the same way as pearlite lamellae are known to be linked? Expressed alternatively, is there one nucleation event per subunit or one per sheaf? The evidence on this is not very clear, but the balance of opinion based on the work of Srinivasan and Wayman (1968), Sandvik (1982) and Sandvik and Wayman (1983) seems to be that each subunit is a separate entity that is nucleated and grows to a limiting size. The reason it stops growing may be because the shape change causes plastic deformation (the austenite being fairly soft at the temperatures of bainite formation) and this destroys the interface coherency.

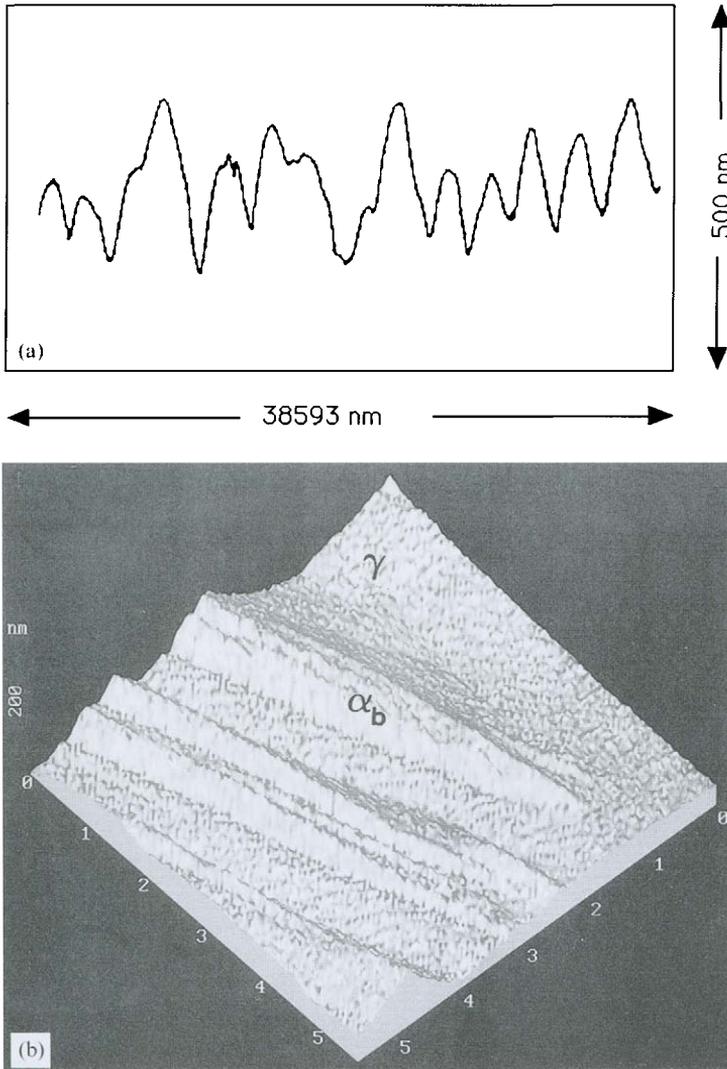


FIG. 25.5. (a) Atom force microscope line scan through basic bainitic ferrite sheaf showing tilt of individual subunit. (b) Atom force microscope image of bainitic ferrite.

Many experiments have shown that there is no partition of any substitutional solutes between the ferrite and the austenite, and this reflects the immobility of the iron and substitutional solutes in the bainitic temperature range. This condition is known as para-equilibrium; at the interface the carbon content is adjusted so that the two phases are in equilibrium subject to the constraint that the substitutional composition is everywhere uniform. It seems obvious that there must be some initial partition of carbon between a

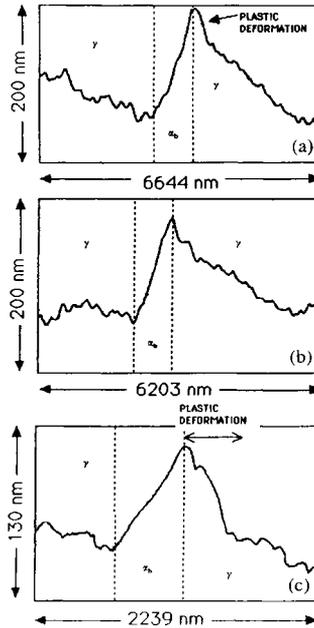


FIG. 25.6. Surface relief due to individual subunits as revealed by atomic force microscope scans. (a) and (b) are from different parts of the same sheaf and (c) is from a different sheaf. Note the plastic deformation needed to accommodate the shape change.

ferrite nucleus and the austenite, because otherwise how could the displacive transformation begin above  $M_s$ ? This leaves the ferrite crystal still supersaturated with carbon, but the remaining excess carbon can diffuse out into the austenite in a very short time (Bhadeshia has estimated it at less than 1 s, in upper bainite). In lower bainite, the carbon concentration is not readjusted so rapidly and the excess carbon trapped in the ferrite precipitates inside the ferrite grains.

Much information about the bainite reaction comes from the “incomplete reaction phenomenon”. It is simply observed that when held at constant temperature the transformation ceases when a certain fraction of bainite has been formed, and the reaction does not continue to completion. In particular, the carbon content of the austenite does not attain its equilibrium value on the  $\gamma/\gamma + \alpha$  boundary of the two phase field, but is limited to the supersaturation at which the austenite and ferrite have the same free energy, i.e. to the curve of equal free energy of the  $\gamma$  and  $\alpha$  phases, allowing for the strain energy. This is the result expected if the basic transformation is displacive, because of course the transformation could not take place if the free energy increased at constant composition.

Granular bainite is a term used to describe the bainite formed during continuous cooling rather than isothermally. It has essentially the same structure as that already described except that the sheaves of bainite tend to be larger and the optical microscope appearance is of blocks of bainite and austenite.

In most alloy steels, the bainite and pearlite modes of decomposition are well separated and there is a bainitic bay between the two C-curves on a  $T-T-T$  diagram. The bainitic reaction is then observed to begin at the particular temperature known as  $B_s$ . In plain carbon steels, however, the two reactions overlap to such an extent that it is only possible to draw one C-curve on a  $T-T-T$  diagram. The upper part of this curve now represents pearlite formation whilst the lower part below the nose represents bainite. When the two reactions are well separated, both a microstructural  $B_s$  and a kinetic  $B_s$  may be defined and are found to agree with each other. Non-lamellar ferrite formed above  $B_s$  is either pro-eutectoid ferrite or divorced pearlite or grain boundary allotropes, and each can be clearly recognized as having a distinctive morphology very different from that of bainite.

This conclusion is not universally accepted. In particular, Aaronson (1969) developed what he termed the microstructural definition of bainite as a non-lamellar aggregate of ferrite and carbide which can form at any temperature below the eutectoidal temperature. The ferrite-austenite interface is supposed to be sessile, and to grow by the migration of ledges (steps) on the interface. The bainite bay is attributed to the effects of solute drag by substitutional solutes which segregate to the interface and, apart from this special effect, the formation of bainite is part of a process which is continuous through the  $B_s$  temperature. In Aaronson's description, no special significance attaches either to  $B_s$  or to the observation of a shape change. This description has led to much controversy in the literature (see especially Aaronson *et al.*, 1970; Hehemann, 1970). There are several objections to the Aaronson model; for example, it does not seem to take account of the lack of any solute atom mobility at the temperatures of bainite formation.

Investigations of the kinetics of bainite formation (Fig. 25.7) have shown that both lengthening and thickening of the ferrite plates and of the sheaves is linear with time. Early measurements of the activation energy for the overall transformation rate in a plain carbon eutectoidal steel showed a transition at a temperature of about 350 °C, the value above this temperature being about twice that below it. The two values were roughly consistent with the energy for carbon diffusion in austenite and ferrite respectively but, when plotted against carbon concentration, both values increased with increasing carbon whilst the activation energy for carbon diffusion decreased with increasing carbon. A more serious discrepancy first arose from measurements by Goodenow *et al.* (1963) of the lengthening rate of individual bainite plates as revealed by surface tilts. They found a reversal of previous findings: the highest activation energy was that of lower bainite and was about twice that for upper bainite. This anomaly was explained by Hawkins and Barford (1972), who measured both overall and individual plate kinetics in both interior and surface bainite. They confirmed the previous results about reversal, but showed that this is due to a variation with temperature of the number of growing bainite plates such that this number is a minimum at about 350 °C. They found the activation energy for nucleation to be roughly constant throughout the whole temperature range and, after a short initial transition, the nucleation rate was balanced by the impingement rate so that, at constant temperature, the number of growing particles remained constant for atomic fractions of solute from 0.2 to 0.6. The authors concluded that it is necessary to study

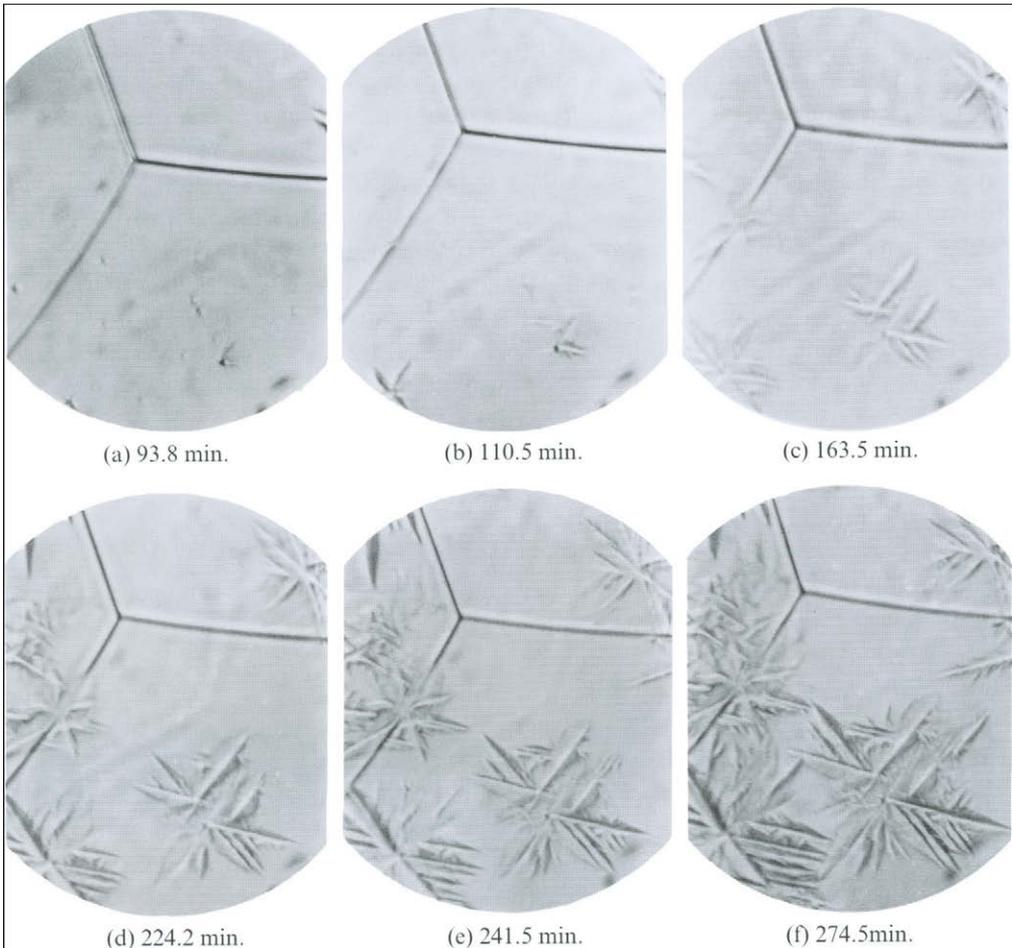


FIG. 25.7. Hot stage micrographs showing the formation of bainite at 265 C in a 1.12 wt.% C 5.28 wt.% Ni iron carbon nickel alloy.

separately the nucleation, growth and impingement factors. When this was done, the surface plate kinetics gave results in excellent agreement with the overall kinetics.

Most measurements show that the subunits and the sheaves grow at a rate which is appreciably faster than that given by carbon diffusion control. This suggests an initial degree of supersaturation which is very rapidly removed in upper bainite but persists until precipitate has formed in lower bainite. Assuming the subunits to be independently nucleated, the overall lengthening rate of a sheaf will clearly be less than that of a subunit and will be dependent on the nucleation rate as well as the growth rate.

The theory of bainite nucleation is not well developed at the present time but, according to a detailed analysis by Bhadeshia (1992), subunits form by a displacive mechanism. Experimental measurements indicate that the energy for nucleation varies

linearly with the driving force, and it is suggested that this results from some form of pre-existing nuclei which grow very slowly until a critical condition is reached. The initial nuclei form almost exclusively on austenite grain boundaries and on any impurity particles which happen to be present but, as each subunit is a separate entity, some form of sympathetic nucleation must occur within the sheaf as it moves forward. Clearly this implies that the growth rate of a sheaf is less than that of a subunit. As already mentioned, several measurements have shown that the growth rate of both sheaves and subunits is too great to be attributed to carbon diffusion control. However, sympathetic nucleation is apparently confined within a sheaf and is not so effective as in martensitic transformation. Several measurements have shown that the rate of growth of both sheaves and subunits is too rapid for it to be controlled by carbon diffusion rates, and it must therefore be concluded that the initial formation of bainite takes place with partial supersaturation.

### 103. CRYSTALLOGRAPHY OF BAINITE

The assumption that the iron atoms and any atoms of substitutional solutes form ferrite in bainite by means of a displacive mechanism implies that the crystallographic theory of Chapter 22 should be applicable to bainite also. The measurements of the change of shape show that the strain energy of a subunit of bainite is comparable with that of some martensitic planes and an unsolved problem is why the habit plane would be different. Greninger and Troiano (1940) first showed that the habit plane is irrational and that it differs from the martensitic habit in the same alloy. In attempting to apply the standard crystallographic theory, one difficulty is that the carbon contents of the coexisting planes are not known and hence neither are the lattice parameters which determine the principal strains of lattice deformation. It is not even known with certainty whether the initial ferrite nucleus has a b.c.c. or b.c.t. structure. It seems certain that there must be some initial partition of carbon between austenite and the ferrite else how could the reaction begin above the  $M_s$  temperature?

The orientation relationship in bainite is irrational but close to the Kurdjumov–Sachs and Nishiyama–Wasserman rational relations. As previously mentioned, both these ideal relations have a set of close-packed planes for each structure mutually parallel. In the Kurdjumov–Sachs relation, close-packed directions within these parallel planes are also parallel, whereas in the Nishiyama–Wasserman relation the close-packed  $\langle 110 \rangle$  direction of the austenite is parallel to a  $\langle 112 \rangle$  direction of the ferrite. The two orientations differ only by a  $5.25^\circ$  rotation about the normal to the parallel close-packed planes and both fall well within the Bain zone of orientation for which the Bain correspondence gives minimum principal strains and no direction or plane normal rotates by more than  $11^\circ$  during the lattice deformation. The fact that the bainite orientation, like that of martensite, always falls within this region is of some significance in showing that there is always a lattice correspondence in the transformation which is, of course, a necessary condition for a displacive transformation. This is established beyond reasonable doubt by the

experimental evidence of the shape deformation and the failure of the ferrite plates to cross grain or twin boundaries.

Most of the early attempts to measure the crystallographic parameters of bainite were applied to the sheaf structure and might thus be considered as some kind of average over the true parameters of the subunits. This is certainly the case in the careful work of Srinivasan and Wayman (1968) which was the first substantial measurement of the relevant detail. They investigated lower bainite and used an alloy of Fe-1.1 wt.% C-7.9 wt.% Cr which had a considerable amount of retained austenite at room temperature. The ferrite particles were found to have only one planar face and this was taken as the habit plane. It was found to be irrational but close to (254) austenite for the variant of the orientation in which (111) was the austenitic plane nearly parallel to a {110} ferritic plane. The habit plane of martensite in the same alloy was close to (494) austenite and, as already emphasized, it is difficult to explain this difference on the basis of the phenomenological theory.

The measured shape change was 0.128 and according to various estimates corresponds to a true shape change of 0.23-0.28 in each subunit. This is clearly compatible with the measurements of Swallow and Bhadeshia and with the crystallographic theory. In comparing their results with the theory, Srinivasan and Wayman showed that the habit plane and orientation relation could be explained by a lattice-invariant shear on an irrational plane in an irrational direction. Alternatively a double shear on the planes (111)<sub>γ</sub> and (101)<sub>γ</sub> in the common direction [101]<sub>γ</sub> will give the observed results if a Bowles-Mackenzie dilatation parameter is incorporated. However, it seems more logical to apply the theory directly to individual subunits.

The first such measurements were made by Sandvik (1982) and by Sandvik and Wayman (1983), who found a habit plane (0.73, 0.663, 0.649)<sub>γ</sub> which is within 6° of the habit found by Srinivasan and Wayman. The orientation relation had an angle of 4° between the close-packed directions, so that it was nearer to Nishiyama-Wasserman than to Kurdjumov-Sachs. Sandvik also measured the shape deformation in a subunit and obtained results in good agreement with other determinations.

Okamoto and Oka (1986) found that, in high carbon steels, there is a mid-rib which on examination in the electron microscope was found to consist of a very thin layer of martensite. The significance of this observation, which may imply the isothermal formation of thin-plate martensite well above the  $M_s$  temperature, has yet to be explored. The effect was not observed below about 1 wt.% carbon.

Investigations of bainite crystallography in a complex iron alloy were made by Hoekstra and his collaborators (Hoekstra, Lelie and Verbraak, 1978; Hoekstra, Ohm and Verbraak, 1978; Hoekstra, 1980). These authors found the orientation relation to be very close to Kurdjumov-Sachs and an irrational habit plane near to (569)<sub>γ</sub>. They concluded that the phenomenological theory cannot be applied to bainite. However, Bhadeshia (1992) has pointed out some internal inconsistencies in their reported crystallography. Other investigations have been made by Shimizu and Tanaka (1978). They used rather large dilation parameters to explain their results, which were based on the hypothesis that the h.c.p. structure was formed as an intermediary.

TABLE XXVI. CHARACTERISTICS FEATURES OF TRANSFORMATIONS IN STEELS  
(BHADESHIA and CHRISTIAN, 1994)

Comment	$\alpha'$	$\alpha_{lb}$	$\alpha_{ub}$	$\alpha_{af}$	$\alpha_{w}$	$\alpha$	$\alpha_i$	P
Nucleation and growth reaction	=	=	=	=	=	=	=	=
Plate morphology	=	=	=	=	=	≠	≠	≠
IPS shape change with shear component	=	=	=	=	=	≠	≠	≠
Diffusionless nucleation	=	≠	≠	≠	≠	≠	≠	≠
Only carbon diffuses during nucleation	≠	=	=	=	=	≠	≠	≠
Reconstructive diffusion during nucleation	≠	≠	≠	≠	≠	=	=	=
Often nucleates intragranularly on defects	=	≠	≠	=	≠	≠	=	≠
Diffusionless growth	=	=	=	=	≠	≠	≠	≠
Reconstructive diffusion during growth	≠	≠	≠	≠	≠	=	=	=
Atomic correspondence (all atoms) during growth	=	=	=	=	≠	≠	≠	≠
Atomic correspondence, during growth for atoms in substitutional sites	=	=	=	=	=	≠	≠	≠
Bulk redistribution of X atoms during growth	≠	≠	≠	≠	≠	⊗	⊗	⊗
Local equilibrium at interface during growth	≠	≠	≠	≠	≠	⊗	⊗	⊗
Local paraequilibrium at interface during growth	≠	≠	≠	≠	=	⊗	⊗	≠
Diffusion of carbon during transformation	≠	≠	≠	≠	=	=	=	=
Carbon diffusion-controlled growth	≠	≠	≠	≠	=	⊗	⊗	⊗
Cooperative growth of ferrite and cementite	≠	≠	≠	≠	≠	≠	≠	=
High dislocation density	=	=	=	=	⊗	≠	≠	≠
Incomplete-reaction phenomenon	≠	=	=	=	≠	≠	≠	≠
Necessity has a glissile interface	=	=	=	=	=	≠	≠	≠
Always has an orientation within the Bain region	=	=	=	=	=	≠	≠	≠
Grows across austenite grain boundaries	≠	≠	≠	≠	≠	=	=	=
High interface mobility at low temperatures	=	=	=	=	=	≠	≠	≠
Displacive transformation mechanism	=	=	=	=	=	≠	≠	≠
Reconstructive transformation mechanism	≠	≠	≠	≠	≠	=	=	=

89.101 A 27

To conclude this brief account of bainite, it is interesting to consider a more general classification of the products of austenite decomposition and our present understanding of the mechanisms involved. Table XXVI (originally due to Bhadeshia) shows how the current theoretical descriptions of how particular transformation products are formed fit in with experimental observations. The products considered are martensite ( $\alpha'$ ), lower bainite ( $\alpha_{lb}$ ), upper bainite ( $\alpha_{ub}$ ), acicular ferrite ( $\alpha_{af}$ ), Widmanstätten ferrite ( $\alpha_w$ ), allotriomorphic ferrite ( $\alpha$ ), idiomorphic ferrite ( $\alpha_i$ ), pearlite (P) and substitutional alloying elements (X). Consistency of an observation with the theoretical description is indicated by “=” and inconsistency by “≠”. In cases where there is sometimes consistency and sometimes not, the situation is indicated by “⊗”. Massive ferrite is not distinguished from  $\alpha$ .

Bhadeshia (1999) considers the unsolved problems in the theory of bainite. He identifies four such problems as follows: (1) the measured growth rate of individual plates of ferrite; (2) a better treatment of autocatalytic nucleation is needed to specify the growth rate of sheaves of ferrite; (3) quantitative modelling of the precipitation of carbide during the

bainite transformation; and (4) features of the transformation of plastically deformed austenite which are not yet understood. The last two issues are the most important from the point of view of practical utilization and Bhadeshia considers that they should be solved first.

## REFERENCES

- AARONSON, H. I. (1962) *The Decomposition of Austenite by Diffusional Processes*, eds. V. F. Zackay and H. I. Aaronson, p. 387, Interscience Publishers, New York; (1969) *The Mechanism of Phase Transformations in Crystalline Solids*, p. 270, The Institute of Metals, London.
- AARONSON, H. I., LAIRD, C., and KINSMAN, K. R. (1970) *Phase Transformations*, p. 313, American Society for Metals, Metals Park, OH.
- AARONSON, H. I., REYNOLD, W. E., SHIFLET, G. T., and SPANOS, H. M. (1990) *Metall. Trans.* **21A**, 210.
- AARONSON, H. I. and WELLS, C. (1956) *Trans. Metall. Soc. AIME* **206**, 1216.
- BHADESHIA, H. K. D. H. (2001) *Bainite in Steels*, 2nd Edition, The Institute of Materials, London; (1999) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-98)*, eds. M. Ahlers, G. Kostorz and M. Sade, *Mater. Sci. Eng. A* **273-275**, 58.
- BHADESHIA, H. K. D. H. and CHRISTIAN, J. W. (1994) *Metall. Trans.* **21A**, 257.
- BHADESHIA, H. K. D. H. and EDMONDS, D. V. (1979) *Metall. Trans.* **10A**, 895; (1980) *Acta Metall.* **28**, 1065.
- CHRISTIAN, J. W. and EDMONDS, D. V. (1984) *Phase Transformations in Ferrous Alloys*, eds. A. R. Marder and J. L. Goldstein, p. 393, The Metals Society of AIME, Warrendale, PA.
- COTTRELL, S. A. and KO, T. (1953) *J. Iron Steel Inst.* **173**, 225.
- GOODENOW, R. H., MATAS, S. J., and HEHEMANN, R. F. (1963) *Trans. Metall. Soc. AIME* **227**, 621.
- GRENINGER, A. B. and TROIANO, A. R. (1940) *Trans. Metall. Soc. AIME* **140**, 307.
- HAWKINS, M. J. and BARFORD, J. (1972) *J. Iron Steel Inst.* **210**, 97.
- HEHEMANN, R. F. (1970) *Phase Transformations*, p. 397, American Society for Metals, Metals Park, OH.
- HILLERT, M. (1962) *Decomposition of Austenite by Diffusional Processes*, eds. V. F. Zackay and H. I. Aaronson, p. 19, Interscience, New York.
- HOEKSTRA, S. (1980) *Acta Metall.* **28**, 507.
- HOEKSTRA, S., LELIE, H. M. M. V. D., and VERBRAAK, C. A. (1978) *Acta Metall.* **26**, 1978.
- HOEKSTRA, S., OHM, R. K., and VERBRAAK, C. A. (1978) *Acta Metall.* **26**, 1505.
- IRVINE, K. J. and PICKERING, F. B. (1963) *J. Iron Steel Inst.* **201**, 518.
- KINSMAN, K. R. and AARONSON, H. I. (1970) *Metall. Trans.* **1**, 1455.
- KO, T. and COTTRELL, S. A. (1952) *J. Iron Steel Inst.* **172**, 307.
- MATAS, S. J. and HEHEMANN, R. F. (1961) *Trans. Metall. Soc. AIME* **221**, 179.
- OBLAK, J. C. and HEHEMANN, R. F. (1967) *Transformations and Hardenability of Steels*, p. 15, Climax Molybdenum Co., Ann Arbor, MI.
- OKAMOTO, H. and OKA, M. (1986) *Metall. Trans.* **17A**, 1113.
- PICKERING, F. B. (1967) *Transformations and Hardenability of Steels*, p. 109, Climax Molybdenum Co., Ann Arbor, MI.
- SANDVIK, B. J. C. (1982) *Metall. Trans.* **13A**, 787, 789.
- SANDVIK, B. J. C. and WAYMAN, C. M. (1983) *Metall. Trans.* **14A**, 809.
- SHIMIZU, K. and TANAKA, Y. (1978) *Trans. Jpn. Inst. Met.* **19**, 685.
- SPEICH, G. R. (1962) *The Decomposition of Austenite by Diffusional Processes*, eds. V. F. Zackay and H. I. Aaronson, p. 353, Interscience, New York.
- SRINIVASAN, C. R. and WAYMAN, C. M. (1968) *Acta Metall.* **16**, 609, 621.
- SWALLOW, E. and BHADESHIA, H. K. D. H. (1996) *Mater. Sci. Technol.* **12**, 121.
- VASUDEVAN, P., GRAHAM, L. W., and AXON, H. J. (1958) *J. Iron Steel Inst.* **190**, 386.

## CHAPTER 26

# *Shape Memory Alloys*

### 104. PSEUDO-ELASTICITY AND HYSTERESIS

A surprising feature of deformation by moving interfaces is that quite large strains produced in this way are often completely reversible when the stress, of either thermal or mechanical origin, is removed again. The shape memory effect, or rather a closely related phenomenon now called pseudo-elasticity, was discovered by the Swedish metallurgist Olander (1932ab) who reported the peculiar rubber-like behaviour of a gold-cadmium alloy at a meeting of the Swedish Metallurgical Society on 27 May 1932. Subsequent investigations of "Olander's alloy" were made by Benedicks (1940) and by Bystrom and Almin (1947).

A proper understanding of the pseudo-elastic phenomena really began with the work of Chang and Read (1951). In addition to a very clear explanation of the rubber-like effect in the martensitic product phase, they made the very important first observations of single interface transformations and hence (albeit for a special case) of the two-way or reversible shape memory effect. They also reported that a straight rod of the cubic phase acquired a permanent set if it was cooled through the transformation under stress. This permanent set disappeared and the specimen straightened again when heated into the cubic phase. This observation is strictly of reverse shape memory under stress, but it comes close to being the first observation of the true shape memory effect.

Following the work of Chang and Read, there were various observations of two pseudo-elastic effects (detailed above) in early papers on copper-zinc (Reynolds and Bever, 1952) and indium-thallium alloys (Burkart and Read, 1953; Basinski and Christian, 1954ab). The term shape memory was not used in any of these papers, but there is at least one reference to an experiment which showed the ordinary shape memory effect. According to Delaey and Thinel (1975), the shape memory effect was first described a little later by Hornbogen and Wassermann (1956) working with copper-zinc alloys. Despite these early papers, strong interest in shape memory alloys did not begin until the observations of the effect in nickel-titanium alloys by Buehler *et al.* (1963) and the detailed studies of Wasilewski (1971).

The experimental phenomena can be classified under four main headings, namely the shape memory effect, the two-way (or reversible) shape memory effect, the superelastic effect and rubber-like (or ferro-) elasticity. The last two effects may also be classified

together as pseudo-elastic effects and are distinguished from the first two by the fact that the deformation is produced only by the application or removal of stress at constant temperature, so that changes in chemical driving force or stress do not occur.

Superelasticity is the formation of martensite under stress at temperatures above  $M_s$ , where the effective stress (or chemical force) is either insufficiently positive or is negative and actually opposes the transformation. Schematic curves showing the superelastic effect are shown in Fig. 26.1. The net deformation after a load-unload cycle is effectively zero. There is a limiting strain which corresponds to the whole specimen being in the product orientation which is most favoured for the type and direction of the applied stress. In Fig. 26.1(a), the nucleation stress is much larger than the growth stress so the deformation begins with a large yield drop, followed by deformation at nearly constant stress. In the reverse deformation on unloading, the interfaces begin to move when the effective stress (chemical plus mechanical) in the reverse direction is sufficient to overcome the resistance to the motion of the interface. This resistance is given approximately by one-half of the stress hysteresis. Figure 26.1(b) shows the form of the stress versus strain curve for the case of easy nucleation where the nucleation stress and growth stress are coincident, whereas the stress needed for nucleation is much larger than that for growth in Fig. 26.1(a).

Experimental superelastic curves with small hysteresis are shown for two different strain rates in Fig. 26.2 and similar plots are given in Fig. 26.3 for the case of difficult

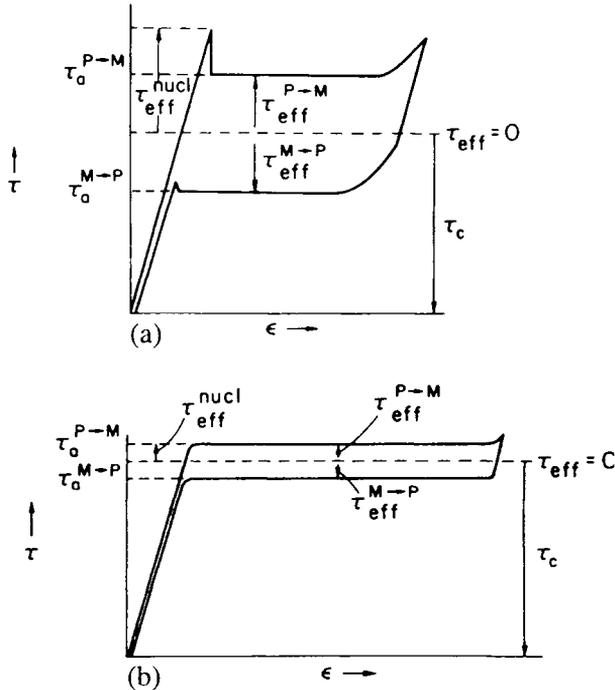


FIG. 26.1. Schematic stress versus strain curves to illustrate superelastic effect.

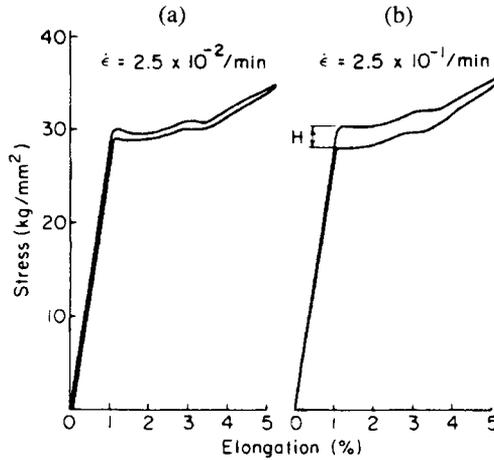


FIG. 26.2. Superelastic effect at two different strain rates for the transformation  $\beta_1 \beta'_1$  in a copper-aluminum-nickel single crystal (after Otsuka *et al.*, 1976).

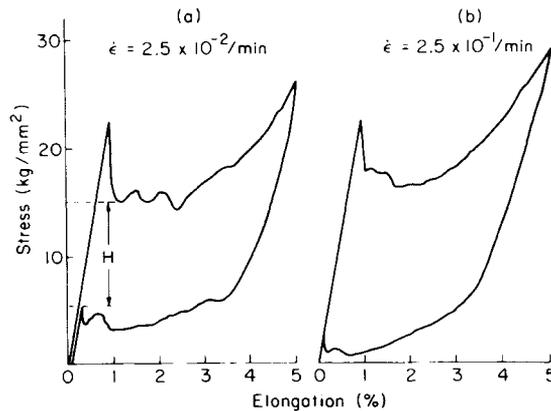


FIG. 26.3. Strain rate dependence of the hysteresis in the  $\beta \beta'_1$  transformation in a copper-aluminum-nickel alloy (after Otsuka *et al.*, 1976).

nucleation and large hysteresis. In both cases, the hysteresis is increased at the higher strain rate, and this is consistent with the assumption that it is attributable to a friction stress which increases with velocity. The interface velocity is a function of the friction stress.

In recent years marked asymmetries have been found in superelastic stress versus strain curves in tension and compression. A difference in the flow stress in tension and compression from different preferred variants with unequal Schmid factors is expected, but (provided the normal component  $\xi$  is negligible) should disappear when resolved shear stresses are compared. Double tension-compression superelastic loops in gold-cadmium

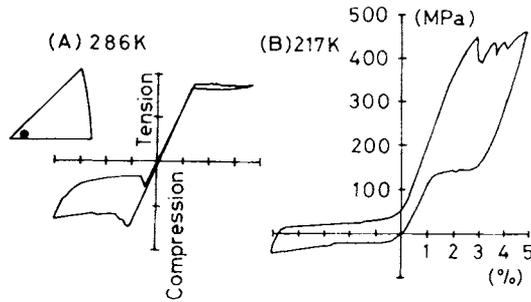


FIG. 26.4. Tension-compression cycle for a single crystal of copper-aluminum-nickel alloy tested (a) above and (b) below  $M_f$ . (After Sakamoto *et al.*, 1979.)

alloys have been attributed simply to variant asymmetry but, in copper-aluminum-nickel alloys, Sakamoto *et al.* (1975) have found marked differences in the shapes as well as the stress levels of the loops. Figure 26.4(a) shows a test above  $M_f$ , i.e. in the stable parent phase. The tensile stresses give rise to the  $\beta_1$ - $\beta'_1$  transformation (cf. Fig. 26.1), while the compressive stresses are relieved by the formation of  $\gamma'_1$  martensite. The authors attribute this difference to non-axiality of the compression load, but there may be a more fundamental explanation. Figure 26.4(b) shows a tension-compression cycle below  $M_f$  where the stress-free phase is  $\gamma'_1$ . The superelastic loop in tension is caused by the  $\gamma'_1$ - $\beta'_1$  transformation, while compression does not produce any phase change but only the displacement of twin boundaries. The strain remains on removal of the stress and disappears only when a small tensile stress is reimposed. Hysteresis loops in which a remanent strain on unloading is removed by opposite stressing have been called "ferroelastic" by Aizu (1969) and by Nakanishi and his co-workers (Nakanishi *et al.*, 1973; Nakanishi, 1975) because of the obvious analogy with ferroelectric and ferromagnetic materials; this is a different usage from that suggested by Lieberman and his co-workers (1970, 1975).

The other kind of superelastic effect is usually called "rubber-like" although it would have been more logical to apply the term to all superelastic events. Rubber-like elasticity is based entirely on the martensite produced in a thermoelastic transformation. If a fully transformed specimen of indium-thallium is extended, bent or twisted to some other shape at room temperature it will comply readily by utilizing the necessary displacements of twin boundaries and interphase boundaries. The same strain applied at low temperatures requires a higher stress, on removal of which the specimen reverts to its original shape by inverse displacements of the grain and twin boundaries.

The effect in gold-cadmium alloys is induced only by annealing of the transformed material, the annealing allowing the alloy to order. Indium-thallium alloys are one of the few thermoelastic or shape memory alloys which do not require ordering, and it has been suggested that this is simply because in the reverse transformation on heating the only possible orientation of the nucleus is that which reproduces the original shape and orientation.

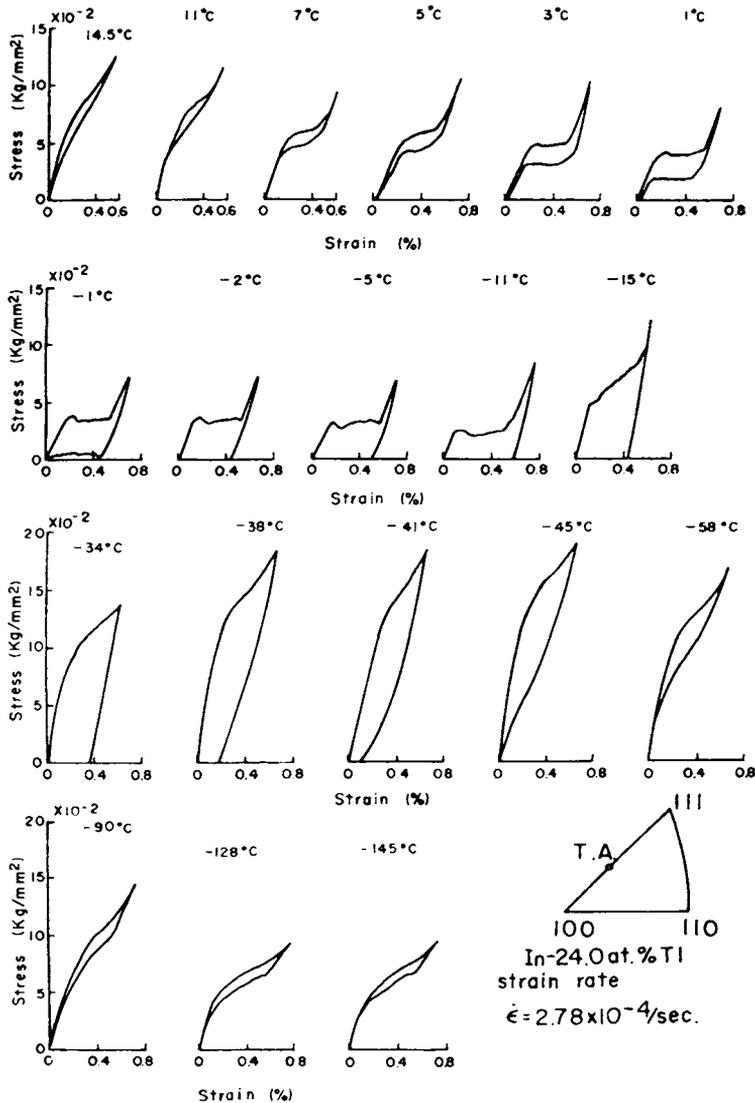


FIG. 26.5. Hysteresis of superelastic, plastic and rubber-like elastic stress versus strain curves for an indium thallium alloy tested at various temperatures.

Figure 26.5 shows a series of stress versus strain curves for an indium–thallium single crystal tested at various different temperatures to illustrate the various effects. At temperatures above  $-25^{\circ}\text{C}$  in this alloy, the superelastic behaviour is produced by the cubic–tetragonal transformation, whilst at lower temperatures down to  $-40^{\circ}\text{C}$ , deformation involves twin boundary migration and leaves a permanent strain. This is the shape memory strain which would be removed by heating back into the f.c.c. region. At still

lower temperatures, below  $-40^{\circ}\text{C}$  for this particular alloy, the strain incorporated by twin boundary displacements, etc. is spontaneously removed on unloading to give the specimen a rubber-like elasticity.

Rubber-like elasticity has been studied mainly in gold-cadmium and indium-thallium alloys, but similar phenomena have been found in copper-zinc (Schroeder and Wayman, 1977), copper-gold-zinc (Miura *et al.*, 1974) and copper-zinc-aluminium alloys (Barcello *et al.*, 1978). The effect is distinct from elastic twinning, in which the accommodation stresses around a twin constitute the driving force for untwinning. Rubber-like elasticity is developed in indium-thallium alloys by simply cooling to a subzero temperature; gold-cadmium alloys have to be given a heat treatment below the  $M_f$  temperature before they become rubber-like. The deformation is produced entirely by the displacement of twin and intervariant boundaries and, on removing the stress, these boundaries simply return to their original configuration. In the case of gold-cadmium, the alloy deforms plastically by twin boundary movements immediately after the transformation, but exhibits rubber-like elastic behaviour when given a short "stabilization" annealing treatment. The rubber-like behaviour is time-dependent; if the applied stress is maintained for a sufficient time "restabilization" occurs in the deformed condition, and rubber-like straining about this new reference point is then possible. Indium-thallium alloys, in contrast, require no initial annealing and in a temperature range below  $M_f$  their behaviour is again perfectly plastic [see Fig. 26.4(b)]. However, restabilization can be obtained by maintaining the stress for a period of time. For both alloys, the results suggest that some extra free energy proportional to the volume swept is generated as a boundary migrates. There has been some speculation about the nature of this force; an early candidate was long-range order because gold-cadmium (although not indium-thallium) is known to be ordered. However, no changes in the long-range order during the annealing have been detected and Ren and Otsuka (1997) have concluded that diffusion of atoms within a sublattice (i.e. a change in short-range order) is responsible for the rubber-like behaviour.

## 105. SHAPE MEMORY

In the ordinary shape memory effect, a specimen which has been transformed to martensite on cooling, but not deformed, is deformed in the martensitic phase and is then reheated into the austenitic phase, whereupon it reverts to its original shape and orientation. The two-way (or "reversible") shape memory effect, in contrast, takes place without any external stress. The specimen deforms spontaneously during the cooling half-cycle and recovers spontaneously during the heating half-cycle.

As a result of many investigations, especially by Wayman and his collaborators (Schroeder and Wayman, 1977; Saburi and Wayman, 1980; Wayman, 1979) and by Otsuka and Shimizu (1970), Tadaki *et al.* (1988) and Murakami *et al.* (1994), there is general agreement about the mechanism of the shape memory effect. Although individual product regions may show tilting of the surface and deviation of fiducial lines, the effect of these shape changes averaged over any macroscopic part of a specimen is near zero. This is

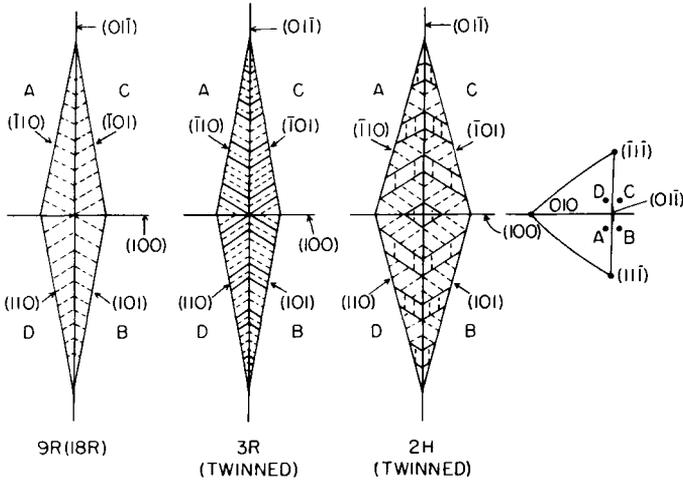


FIG. 26.6. Possible arrangement of variants around a  $(110)$  pole to form a self-accommodating group in various copper alloys (after Wayman, 1979).

because the alloys concerned are very efficient at self-accommodation of the individually transformed volumes. A possible arrangement is for groups of four different variants to group together in a diamond-shaped pattern around a  $\{110\}$  type plane which is at a comparatively small angle to the various habit planes. The transformation is then completed by the formation of intervariant interfaces, as shown in Fig. 26.6. A rather more probable arrangement of four variants is in a parallelogram morphology as shown in Fig. 26.7 (Murakami *et al.*, 1994). The four variants  $A$ ,  $B$ ,  $C$  and  $D$  are such that  $A$  and  $B$  and  $C$  and  $D$  are in type II twinning relation to each other whereas  $A$  and  $C$  and  $B$  and  $D$  are type I twins of each other. Thus all the interfaces are highly mobile.

The fully martensitic state of such an alloy thus consists of a number of differently orientated regions, separated by intervariant interfaces and by twin interfaces (see Fig. 26.8). The common characteristic of all the interfaces is that they are still glissile and may be displaced by an applied stress or a chemical driving force. Suppose the specimen was originally a single crystal of the parent (high temperature) phase. Deformation of the martensitic structure then proceeds by twin boundary migration within each variant and by the growth of favourably orientated variants with respect to the applied stress. The details of the variant conversion sequence depend on the particular alloy and the structure of its martensite, but typically there are two stages. The four variants in each group are first converted to the single most favoured variant in that group whilst, in the second stage, intergroup conversion produces a single preferred variant throughout the specimen, which (provided the ductility is adequate) is eventually converted into a martensitic single crystal. Occasionally the most favoured variant grows directly into the four variants of an adjacent group. The structure, orientation and shape change from the original cubic crystal is now equivalent to that of stress-induced martensite at the limiting strain of a superelastic loop,

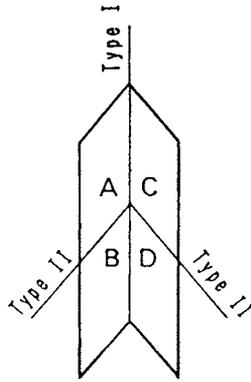


FIG. 26.7. Parallelogram arrangement of variants to form a self-accommodating group (after Murakami *et al.*, 1994).

apart from trivial changes arising from the difference in temperature of the two operations. However, on removing the stress, there is no further change of structure or dimensions and the final reheating causes the single crystal of martensite to revert to a single crystal of the parent phase. If this single crystal has the same dimensions as the original specimen, the shape change produced by the low temperature deformation must be exactly cancelled during the high temperature phase change.

Clearly in a polycrystalline specimen not every grain will be able to change completely into a single crystal because of compatibility problems, but the general principles of the above description still apply. Another factor is that not all possible variants may be formed during the initial phase change, so the deformation may be restricted to the growth of the most favourably orientated variant among those present. In the extreme case, a single interface transformation may still give a shape memory specimen as deformation can occur by migration of the transformation twin boundaries. Such a specimen changes shape both during the cooling transformation and again when stressed in the transformed state and the shape change which is removed on heating is the net effect of these two changes. The whole cycle may thus be regarded as the combination of the reversible shape memory and the normal shape memory processes, and this also applies to any specimen which undergoes some macroscopic shape change on simple cooling because of the formation of preferred variants.

Simple extension or compression will ideally lead to a single crystal but more complex shape changes involving bending or twisting can be obtained by a suitable distribution of twin-related crystals. An extension of this argument shows that virtually any shape may be obtained by variant conversion, provided only that the strains nowhere exceed the maximum permitted by the parameters of the macroscopic shape deformation **E**. Thus the shape memory effect is not limited to "simple" deformations.

In polycrystalline specimens, the additional constraints will increase the stresses and will considerably reduce the maximum strain, as in superelastic deformation, but if each grain regains its original orientation during the heating transformation, there will again be a full

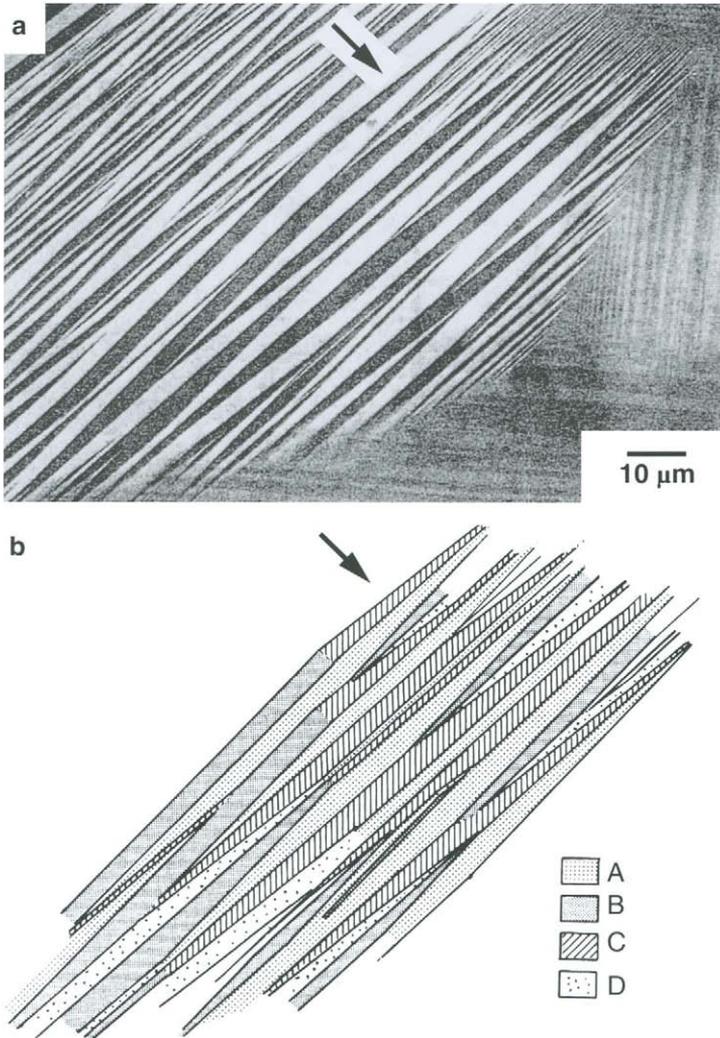


FIG. 26.8. (a) Typical micrograph of self-accommodating variants. (b) Variants identified from trace analysis, corresponding to those shown in Fig. 26.8 (after Murakami *et al.*, 1994).

shape memory effect. The cycle of changes has been illustrated diagrammatically by Saburi *et al.* (1979); Fig. 26.9 shows a similar summary due to Cook (1981), who compares the changes involved in a shape memory cycle  $VABCDEFZ$  with those in a superelastic cycle  $VWXYZ$ . Provided the same martensite phase is produced in the two cycles (an important qualification in many alloys) the structure at  $X$  is identical with that at  $C$  and  $D$  and state  $X$  may be converted into  $D$  by cooling at constant strain (and varying applied stress) as indicated by the arrow.

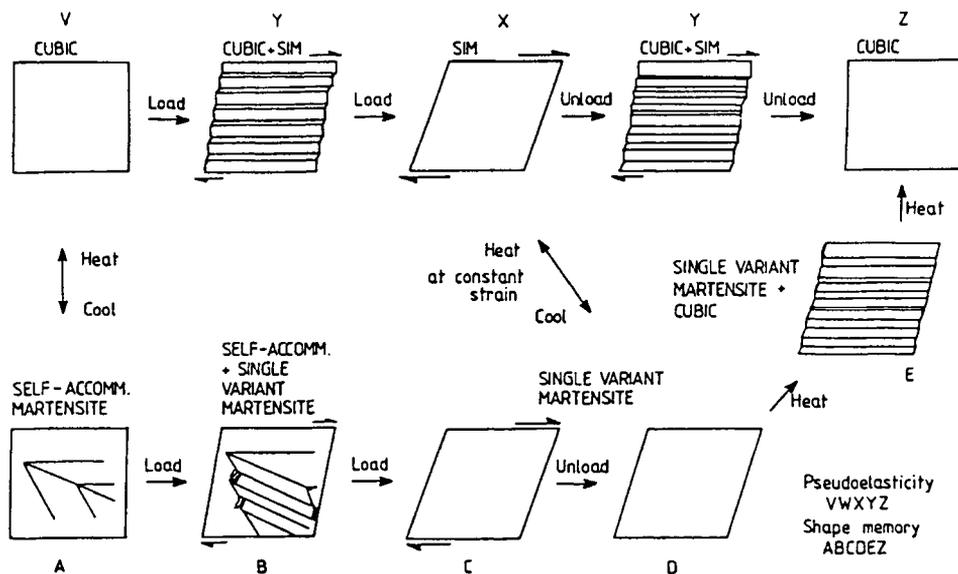


FIG. 26.9. Schematic structural changes during a superelastic cycle  $VWXYZ$  and a shape memory cycle  $VABCDEZ$ .

Cook tested the equivalence for polycrystalline specimens of copper-zinc-aluminum which in separate experiments were taken to state  $D$  via routes  $VABCD$  and  $VWXD$ . Each specimen was then loaded in tension to various levels and the temperatures at which the reverse (heating) transformation began were measured. This gives a Clausius-Clapeyron type plot in which  $\Delta g_{\text{mech}}$  opposes the transformation and thus raises the transformation temperature. The results of a plot of temperature versus opposing stress are straight lines of the same slope for specimens produced by the two routes, thus indicating that the initial states were identical.

The above description is rather macroscopic and much remains to be understood in terms of the deformation and reversion atomic scale mechanisms. Most shape memory alloys are ordered (indium-thallium is a notable exception) and many are internally twinned or faulted. Ordering helps to promote thermoelastic deformation by reducing the probability that accommodation stresses will cause the plastic deformation. Reversibility in the shape memory cycle of Fig. 26.9 is obtained only if the last stage  $E \rightarrow Z$  is identical with the stage  $W \rightarrow V$  and hence is exactly the inverse of  $V \rightarrow W$ . Hence the parent phase must form plates with the same variant of the reverse correspondence as in the reversion of stress-induced martensite. Clearly this is crystallographically possible in all cases; a main feature of shape memory alloys is believed to be that this is the *only* possibility. In many transformations, the symmetry of the martensite is low enough to ensure that there is only one possible reverse path back to the parent phase. In particular, a lower symmetry is obtained if the phases are ordered, as alternative correspondences for the reverse transformation, which would simply be variants in a disordered structure, become

non-equivalent and lead to the disruption of the long-range order. An additional factor arises if the lattice-invariant shear is produced by twinning or by partial dislocations which leave stacking faults in the martensite. The shear plane is then distinguishable from the other lattice planes of that crystallographic form; i.e. the macroscopic symmetry of the martensitic phase has been lowered, and it is this symmetry rather than that of a unit cell which is relevant to the inverse transformation.

Even with a unique martensite–parent correspondence, there are two possible habit planes each giving the same shape deformation but with a rotation of the final parent lattice, as described in Section 2. In an unconstrained single crystal this could lead to a net rotation of the parent lattice in a cycle in which the preferred variant forms on one invariant plane and then retransforms on heating on the conjugate invariant plane. This does not often happen, although it has been observed in single interface transformations, and Cook suggested that the constraints of the grips in a testing machine prevent the rotation and thus ensure complete reversibility.

The reversible (or two-way) shape memory effect is due to the formation of one variant of the product phase preferentially during cooling. This can be achieved by nucleation and growth of only one variant, as in a single interface experiment, but, in most specimens, the bias probably originates in non-random internal stresses due to plastic accommodation of plates formed in a previous cycle. Many experiments have shown that the reversible shape memory is developed by a “training” or “shakedown” treatment of repeated cycling around either a shape memory or a superelastic sequence, and the strains which develop on cooling are then always in the same sense as those produced by stress in the appropriate part of the superelastic cycle. Hence in shape memory cycles, the permanent deformation produced by straining the product phase diminishes in magnitude as the strain obtained on cooling increases, and the sum of these two remains approximately constant.

#### REFERENCES

- AIZU, K. (1969) *J.Phys. Soc. Jpn.* **27**, 387.  
 BARCELLO, G., TACCI, W., and AHLERS, M. (1978) *Scr. Metall.* **10**, 1075.  
 BASINSKI, Z. S. and CHRISTIAN, J. W. (1954a) *Acta Metall.* **2**, 101; (1954b) *Ibid.* **2**, 148.  
 BENEDICKS, C. (1940) *Ark Mat., Astron. Fys.* **27A**, 15.  
 BUEHLER, W. J., GIEBRICH, J. V., and WILRY, R. C. (1963) *J. Appl. Phys.* **34**, 1475.  
 BURKART, M. W. and READ, T. A. (1953) *Trans. Metall. Soc. AIME* **197**, 506.  
 BYSTROM, A. and ALMIN, K. E. (1947) *Acta Chem. Scand.* **1**, 76.  
 CHANG, L. C. and READ, T. A. (1951) *Trans. Metall. Soc. AIME* **191**, 47.  
 COOK, J. M. (1981) *Ph.D. Thesis*, Cambridge University, Cambridge, UK.  
 DELAËY, L. and THINEL, J. (1975) *Shape Memory Effects in Alloys*, ed. J. Perkins, p. 341, Plenum Press, New York.  
 HORNBOKEN, E. and WASSERMANN, G. (1956) *Z. Metallkd.* **47**, 427.  
 LIEBERMAN, D. S. (1970) *Phase Transformations*, p. 1, American Society for Metals, Metals Park, OH.  
 LIEBERMAN, D. S., SCHMERLING, M. A., and KATZ, R. W. (1975) *Shape Memory Effects in Alloys*, ed. J. Perkins, p. 203, Plenum Press, New York.  
 MIURA, S., MAEDA, S., and NAKANISHI, N. (1974) *Philos. Mag.* **30**, 565.  
 MURAKAMI, Y., OTSUKA, K., HANADA, S., and WATANABE, S. (1994) *Mater. Sci. Eng. A* **189**, 191.  
 NAKANISHI, N. (1975) *Shape Memory Effects in Alloys*, ed. J. Perkins, p. 147, Plenum Press, New York.

- NAKANISHI, N., MORI, T., MIURA, S., MURUKAMI, Y., and KACHI, S. (1973) *Philos. Mag.* **28**, 277.
- OLANDER, A. (1932a) *Z. Kristallogr., Kristallgeom., Kristallphys., Kristallchem.* **83A**, 145; (1932b) *J. Am. Chem. Soc.* **54**, 3819.
- OTSUKA, K. and SHIMIZU, K. (1970) *Scr. Metall.* **4**, 467.
- OTSUKA, K., WAYMAN, C. M., NAKAI, K., SAKAMOTO, H., and SHIMIZU, K. (1976) *Acta Metall.* **24**, 207.
- REN, X. and OTSUKA, K. (1997) *Nature* **389**, 579.
- REYNOLDS, J. E. and BEVER, M. B. (1952) *Trans. Metall. Soc. AIME* **194**, 1065.
- SABURI, T., NENNO, S., and WAYMAN, C. M. (1979) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-79)*, p. 619, MIT Press, Cambridge, MA.
- SABURI, T. and WAYMAN, C. M. (1980) *Acta Metall.* **28**, 1.
- SAKAMOTO, H., TANAGAWA, M., OTSUKA, K., and SHIMIZU, K. (1975) *Proc. Int. Conf. Martensitic Transformations (ICOMAT-75)*, p. 633, MIT Press, Cambridge, MA.
- SCHROEDER, T. A. and WAYMAN, C. M. (1977) *Acta Metall.* **25**, 1375.
- TADAKI, T., OTSUKA, K., and SHIMIZU, K. (1988) *Annu. Rev. Mater. Sci.* **18**, 25.
- WASILEWSKI, R. J. (1971) *Metall. Trans.* **3A**, 2973.
- WAYMAN, C. M. (1979) *Spring Conf. Inst. Metallurgists*, p. IV-1, Institute of Metallurgists, London.

# Subject Index

Alloy systems are indexed alphabetically, irrespective of major and minor components

- Absolute activity 179
- Accommodation deformation 463
- Accommodation factor 464, 469–71
- Activated complex 85–91, 137, 407
- Activation energy (enthalpy) 89–90
  - experimental 83–4, 90, 408–13, 486
  - for boundary migration 485–6
  - for defect motion 137–9, 409–12
  - for diffusion 136, 380–1, 408–13, 442
  - for grain boundary diffusion 380–1, 413–14, 416
- Activation entropy 89–90
  - for defect motion 137–8, 409, 411, 413
  - for diffusion 408–11, 413
- Activation free energy 89–90
  - for boundary migration 443–5, 452, 482–5
  - for defect motion 137–9, 410
  - for diffusion 381, 407–8, 410
  - for dislocation diffusion 416
  - for grain boundary diffusion 413–14, 416
  - Zener formulation of 89–90, 408, 410
- Activities 88
- Activity coefficient 88–90, 179, 182, 191, 197
- Additive reactions 549–51
- Adsorbed atoms 149–50
- Adsorbed impurities 272–3
- Affine transformation 43
- Affinity of reaction 102–3
- Alkali metals 117, 380
- Allotriomorphic growth 327
- Allotropic transformations 107–11
- Aluminium 76, 135–7, 298, 300, 305, 308, 315
  - copper alloys 210
  - iron alloys 210
  - zinc alloys 185, 208
- Amplification factor 399–400
- Anisotropic growth 529
- Anisotropy
  - elastic 75, 117, 262, 267–8, 305, 308, 313, 466, 470, 472
  - factor 75–6, 117
  - of diffusion 381, 387
  - of physical properties 104
- Annihilation of opposite defects 140–1, 241
- Antiferromagnetism 107, 109
- Antimony 161
- Antiphase domains 215–16, 228
- Arrhenius
  - equation 83–4, 412
  - plot 136, 143, 145, 410–13, 416, 418, 547
- Athermal nucleation 10–11
- Athermal transformations 13–14
- Atomic domain 120
- Atomic mobility 392–9
- Atomic volume 172, 184–5, 202–5, 269–72, 398, 424–7, 464
- Atomistic calculations 107–22, 125–7, 130–4, 137–9, 163, 165, 167, 277–9, 310–13, 350, 352–4, 356
- Average potential model 199
- Avrami equation 20–1, 532, 541, 544, 546
- Axes
  - Cartesian 24
  - curvilinear 36–7
  - orthonormal 24, 53, 59
  - principal 43, 46–50
  - reciprocal 27
  - transformation laws 31–7
- Axial ratio 120–2, 209
- Bad crystal 106, 244, 246, 287
- Bardeen–Herring source 250, 252, 403
- Basis of structure 25–6
- Beltrami–Michell compatibility equations 77
- Bethe theory of ordering 154–5, 218, 220–4
- Binding energy 1, 111, 161
  - of solute atom to dislocation 269–72
  - of solute atom to vacancy 404
  - of vacancy cluster 221
  - of vacancy pair 130–1
  - of vacancy to dislocations 238–9
- Bismuth 161
  - magnesium alloys 198
- Bloch
  - function 113

- Bloch (*continued*)  
   wall 166  
   Body-centred cubic structure 38-9, 107, 109-11, 113, 116-21, 125-7, 134, 138, 270-2, 310-13, 382, 386, 409, 411-13  
 Body forces 68  
 Bohr magneton numbers 109  
 Boltzmann  
   equation 220, 547  
   factor 231  
   statistics 99  
   substitution 388-90, 494  
 Bordoni peak 280  
 Born-Mayer  
   potential 131  
   repulsive forces 277  
 Bragg-Williams theory 155, 217-27  
 Branching 509-27  
 $\alpha$ -Brass 382  
 $\beta$ -Brass 76, 117, 210, 219  
 Bravais lattice 24-6, 107, 173, 209  
 Brillouin zones 111, 214  
 Bubble model 338  
 Bulk modulus 77, 203-6  
 Burgers  
   circuit 244-6, 287, 290, 300, 317-19, 342, 367, 370, 376  
   vector 236, 238, 240-6, 248-50, 252-3, 255, 258, 260, 262-6, 269, 274-5, 286-301, 303-4, 307-11, 313, 316-21, 336-7, 339, 341-5, 349, 351-2, 360-1, 363-7, 370, 372-6, 476-7  
  
 Cadmium 309, 381  
   -magnesium alloys 210  
   -silver alloys 210  
 Caesium chloride ( $L2_0$ ) structure 210  
 Cauchy relations 74, 111, 115  
 Cavities as nucleation sites 454-6  
 C curves 547-8  
 Cell factors 38-9, 41  
 Cellular growth 508-27  
 Central forces 74, 111-22  
 Centrosymmetric structures 26, 51  
 Cerium 413  
 Change of axes 31-7, 43-5  
 Characteristic equation 44  
 Chemical  
   defects 127  
   diffusion 378, 383-5, 397-8, 404-7  
   diffusion coefficient 383-4, 394  
   potential 175-7, 179, 182, 190, 193, 207, 391-2, 394, 398  
 Chromium 110, 118, 413  
  
 Civilian transformations 6  
 Classification of transformations 1-16  
 Close-packed structures 110-11, 120-4, 127, 162-4  
 Close-packed surfaces 148, 152-5  
 Cluster variation method 196, 224-5  
 Clustering 215, 228-33  
 Cobalt 122, 125, 298, 309  
   -gold alloys 208  
   -iron alloys 210  
   -platinum alloys 210  
 Coefficients of connection 321  
 Coherent boundaries 353, 359-63, 366, 370-1  
 Coherent particles 461-2, 466, 469-71, 476-7  
 Coincidence site boundaries 346, 348-54, 485  
 Coincidence site lattice 346-53, 366, 368, 373-4  
 Cold work  
   effect on diffusion 418-19  
   effect on nucleation 463  
 Collision frequency 90  
 Columbium *see* Niobium  
 Compatibility conditions 68  
   Beltrami-Michell form 77  
 Compatible deformations 316, 319-21, 362  
 Complete pattern shift lattice 374  
 Compliances 73-5  
 Composition fluctuations 184-5, 228-33  
 Concentration contours 390, 414-15, 488, 510-11  
 Concentration parameters for growth 491-4, 499-505, 515, 518, 520-2  
 Conformal solutions 198-9  
 Contact angle 453, 455-9  
 Continuity of state 166  
 Continuous distribution of dislocations 274-5, 316-22, 365, 370  
 Contravariance 35-7  
 Co-operative phenomena 153, 218  
 Coordinates  
   Cartesian 24  
   curvilinear 36-7  
   cylindrical 259-60  
   orthonormal 24  
   polar 202-4  
 Coordination number 120-1, 188, 195-6, 211  
   in liquids 161-2  
 Copper 76, 117-18, 130-1, 133, 135-6, 145, 147, 298-9, 338, 382  
   -gold alloys 210, 226-8  
   -palladium alloys 210  
   -platinum alloys 210  
   -silver alloys 208  
   -zinc alloys 76, 117, 210, 219, 382  
 Coriolis force 98  
 Correlation effects in diffusion 396-7, 404-7  
 Cottrell atmosphere 272-3, 507, 545

- Covariance 35–7  
 Critical nucleus size 426–8, 431, 441–3  
 Critical radius of dislocation loop 281  
 Critical resolved shear stress law 235, 256  
 Critical temperature 153–5, 190, 194–7, 207–8, 218–19, 221–5  
 Cross-slip 250, 253, 315  
 Crowdion 128, 133–4, 380  
 Crum's theorem 469  
 Crystal  
   bad and good 106, 244, 287  
   close-packing 162–4  
   ideal 23  
   macroscopic concept 23  
   nearly perfect 314  
   reference 244–5, 290–1, 317–22  
 Curie temperature 108–10, 153  
 Cybotaxis 162
- Darken's relations 394–5  
 Debye  
   frequency 92, 426  
   model of solid 80, 107  
   temperature 93–4, 107–10, 223, 411  
 Deformation  
   dislocation 322  
   faults 124  
   general 43–4  
   homogeneous 42–50, 63–6  
   infinitesimal 63–8  
   inhomogeneous 66–8  
   lattice 56–7, 62–3, 322, 365–76  
   lattice invariant 322  
   matrix representation 42–3  
   shape 321–2  
   tensor 42  
 Degree of advancement 102  
 Degree of segregation 518–19  
 Degree of supersaturation 427, 435  
 Dendritic growth 495, 498, 501–4  
 Density 135, 166  
 Depleted zone 515, 518  
 Detailed balancing, principle of 83, 98, 150, 432  
 Determinant 32, 35, 44, 46, 48–9  
 Diffuse interface 5, 166–7, 481–2  
 Diffusion  
   activation energy 136, 381, 407–13, 442  
   along cellular boundaries 517–27  
   along dislocations 416–18  
   along grain boundaries 413–18  
   effects of plastic deformation 419  
   excess vacancy concentrations 418–19  
   in non-cubic crystals 381, 387  
   isotope effect 92–4, 408–9, 411  
   mechanisms 378–82  
   phenomenological theory 391–7  
   short circuits 410  
   uphill 384, 397–400  
 Diffusion coefficient 383–7, 390–1, 406–13  
   chemical 383–4, 394  
   concentration dependence 390, 493–5  
   dislocation 416–18  
   experimental measurement 388, 390–1, 414–16, 418  
   for vacancies 395  
   grain boundary 413–18  
   intrinsic 383–4, 393–5  
   self 383–4, 396  
   temperature variation 135–6, 409–13  
   theory of 407–13  
   tracer 383–4, 394–5  
 Diffusion-controlled growth  
   of cylinders 493  
   of dendrites 495, 498, 501–4  
   of duplex regions 508–27  
   of ellipsoids 490, 494–5  
   of plates and needles 498–505  
   of spheres 490, 492  
   stability of 495–8  
   steady-state solution 490, 492–5, 497–8, 500–2, 504–5, 514–17  
   Zener–Hillert model 498–505  
 Diffusionless transformations *see* Martensitic transformations  
 Dihedral angles 331–2  
 Dilatation 61, 68, 203–6, 466, 469–71, 476–8  
 Dilute solutions 180–3, 189, 404–6  
 Direct neighbours 121, 163  
 Direction indices 27  
 Dirichlet  
   cell 121  
   region 120  
 Disclinations 257–8, 361  
 Discontinuous precipitation 508, 519–21, 523, 525–7  
 Discriminating cubic equation 47  
 Dislocation  
   boundaries 334–76  
   core 257, 261–2, 273–8, 311–13  
   deformation 322  
   networks 315–16  
   nodes 246, 249–50, 291–4, 304, 306–7  
   reactions 246, 295, 297, 299–300, 302, 304, 306, 309–11  
   width 237, 274–8  
 Dislocation density  
   in crystals 248, 313–16  
   in grain boundaries 336–8, 342–5  
   in interphase boundaries 365–76  
   tensor 316–20

## Dislocations

- Burgers vector of 236, 238, 240–6, 248–50, 252–3, 255, 258, 260, 262–6, 269, 274–5, 286–301, 303–4, 307–11, 313, 316–21, 336–7, 339, 341–5, 349, 351–2, 360–1, 363–7, 370, 372–6, 476–7
- climb of 238–9, 250–3, 255, 288–9, 363–5, 403–4
- diffusion along 416–18
- edge type 236–45, 257–62, 264–5, 271–2, 274, 277–9, 281, 283–4, 289–90, 297–9, 336–7, 339, 345, 363–4
- elastic energy of 261–7, 275, 294
- energy of loop 265–7, 281
- extended 287, 298–302, 304–5, 308–10
- force on 254–6, 264–6
- Frank partial 296, 300, 304–7, 310
- free energy of 263
- glide of 236, 238–9, 241, 246–54, 287–9, 291–2, 294, 365
- helical 251–2
- imperfect (partial) 286–8, 295–311
- in anisotropic materials 262, 267–8, 305, 308, 313
- in b.c.c. metals 310–13
- in elastic continuum 257
- in f.c.c. metals 295–9, 303, 306, 308–10
- in h.c.p. metals 308–10
- interaction energy 263–7
- interaction with solutes 269–73
- jogs in 238–9, 252–4, 301, 304–5, 307
- kinetic energy of 281, 283–4
- kinks in 279–80
- line tension of 243, 267–8
- nucleation of 280–1
- nucleation on 472–8
- Peierls–Nabarro model 273–81, 284
- prismatic 244, 250, 252–3, 266, 300
- screw type 236, 239–45, 252–4, 259–60, 262, 265, 269–72, 276–80, 282–4, 289–90, 292–3, 298–9, 311–13, 339, 341
- sessile 288, 296–7, 300–2, 309–10
- Shockley partial 296, 300, 302, 305–6, 308, 310
- Somigliana 257–8
- sources of 248–52
- stair-rod 301–3, 305–7
- stress field of 258–61
- surface 322, 365–6, 376
- transformation 291–3
- twinning 288–94
- velocity of 282–6
- Volterra 257–8, 260
- Displacements 63–8, 77, 202–3, 466–7
- Displacive transformations 6
- Distribution function for embryos 429–30, 433–4, 436–8, 444, 446–51

## Divacancy

- binding energy 130–1, 143, 145
- contribution to diffusion 410–13
- equilibrium concentration 130, 137
- free energy of formation 143–5
- free energy of motion 143–4
- in quenched metals 141–6
- orientation of 130–2
- Doppler broadening 136
- Double lattice structures 26
- Drift flow 140, 383, 506–8, 545
- Driving force 3–4, 398–9, 481–6, 501–3, 511–13, 521, 527
- DSC lattice 374–6
- Dynamical theory of rate processes 81–94
- Eigenvalues of matrix 44, 47
- Eigenvectors 44–5, 47
- Einstein
  - equation 395
  - frequency 92
  - model of solid 80
  - Smoluchowski equation 385, 411
- Elastic anisotropy factor 75–6, 117
- Elastic constants 73, 75–7
- Elastic moduli 75–7
- Elasticity, theory of 68–77
  - Electrical resistance measurements 141–3, 146, 227
- Electron microscopy 142, 146, 228, 248, 253, 280, 299, 302, 305–6, 308, 354, 374
- Electron theory of metals 111–13
- Elongation quadric 50, 66
- Embryos 425–44, 446–59
- Energy cusps 359–60
- Engineering strains 72
- Entropy
  - configurational 128–9, 174, 178, 187–8
  - excess 180–1, 194, 196–9, 201, 206
  - of activation 89–90, 92, 137–9, 220, 408–11, 413
  - of fusion 446
  - of mixing 173–4, 178, 180–1, 187–8, 194, 196–9, 201, 206
  - of solution 180–1
  - partial 180–1
  - source strength 101–2, 483
  - thermal 174, 189, 191, 198
- Equations
  - of compatibility 68, 77
  - of equilibrium 71, 77, 202
  - of motion 71, 77, 282
  - of state 118, 161
- Equilibrium 1, 10, 81–3, 88–9, 95–6, 98–101, 104
  - constant 82–3, 88

- dynamic 82, 432
- of three-dimensional grains 333–4
- of three-grain junction 331–2
- of two-dimensional array 330–1, 333
- shape 155–60, 441–2, 463–6, 468–72
- unstable 427
- Equivalent lattice points 363
- Eutectic alloys 177, 180, 190
- Eutectoidal growth 508–19, 521–7
- Evaporation energy 1, 149, 151
- Excess entropy of mixing 180–1, 194, 196–9, 201, 206
- Excess free energy of mixing 180–1
- Exchange sites 149
- Experimental activation energy 83–4, 89–90, 408–13, 486
- Extended area 530–1
- Extended dislocation 287, 298–302, 304–5, 308–10
- Extended volume 18–19, 531
- Extensions 66
- Extrinsic faults 123–4, 296–300, 302, 304–5, 309–10
  
- Face-centred cubic structure 39, 107–8, 110, 113, 116–18, 120–8, 131–2, 295–9, 303, 306, 308–10, 362, 409–13
- Fermi
  - Dirac statistics 112
  - distribution law 273
  - energy 113, 117, 122
  - surface 2, 111–13, 119
- Ferromagnetism 107–10, 166, 218
- Fibre structure 327
- Fick's Law 378, 383, 385, 387, 391, 408
  - differential form 387
  - modification of 399–400
- Field-ionmicroscopy 148, 351
- Finnis–Sinclair potential 2, 113
- First ellipsoid 48
- First order phase changes 161, 225–7
- First order reaction 16, 21, 81, 140
- Fluctuations 4, 98–9, 228–33, 422, 429–31, 433
- Foam structure 216
- Fokker–Planck equation 449
- Force between two dislocations 264–6
- Force on dislocation 254–6
- Frank partial dislocation 296, 300, 304–7, 310
- Frank–Read source 248–52, 268, 463, 472
- Free energy
  - driving force 3–4
  - excess 180–1
  - of critical nucleus 427–8, 443, 452–3, 456–61, 475–8
  - of dislocation 263
  - of divacancy 143–5
  - of grain boundary 166–7, 328–9, 356–61
  - of interphase boundary 365
  - of interstitial 133–4
  - of liquid–solid interface 166
  - of mixing (orsolution) 174–81, 188–9, 193–4, 197, 199, 206–7
  - of ordered alloy 218
  - of stacking fault 125–7, 298–300, 304–11
  - of vacancy 128–32, 135–6
  - partial 175
    - composition curves 174–9, 188–9
- Free volume 164
- Frenkel kinks 146, 149–51
- Full structures 111, 127
  
- Gallium 161
- Generating deformations 317–21
- Generating node 293–4
- Germanium 76, 161, 314, 413
  - manganese alloys 210
- Gibbs
  - Duhem equation 393–4, 398
  - free energy 4, 200
  - function 101, 107
  - Thomson effect 182–4, 427, 487, 490, 495–9, 501–4, 512, 516, 523, 526
- Glide plane 235–8, 240–3, 245–50, 252–3
- Glide surface 246–8
- Glide twinning 52–6
- Glissile boundaries 6–7, 365
- Gold 76, 113, 135–6, 141–5, 307, 411
  - iron alloys 208
  - nickel alloys 185, 198, 206, 208, 229
  - platinum alloys 208
- Gradient energy 185, 398, 400
- Grain boundaries
  - bubble model 338
  - coincidence site 346, 348–54, 485
  - dislocation models 334–76
  - energy of 166–7, 328–9, 356–61
  - general 328, 341–4, 360
  - high angle 338, 344–6, 349, 351, 353–5, 360–1, 414, 416
  - low angle 336–8, 341, 343–5, 351–6, 360–1, 414, 416–17
  - tilt 335–9, 351–2, 354–6
  - twist 340–1
- Grain boundary
  - area per unit volume 334, 532–5
  - diffusion 413–18
  - mobility 484–7, 515, 519, 521–3
  - nucleation 456–61
- Grain corners per unit volume 334, 532, 536

- Grain edge length per unit volume 334, 532, 536  
 Grain growth 330-1, 333-4, 485-6  
 Ground states 211-14  
 Growth faults 124  
 Growth laws  
   linear 16, 481-6, 498-505, 508-27  
   parabolic 18, 486-95  
 Growth parameters 491-4, 499-505, 515-16, 518, 520-2  
 Growth rate  
   dependence on driving force 482-6, 492-5, 497-8, 500-2, 504-5  
   maximum value condition 499, 501, 504, 513, 516-19, 525  
   of curved interface 483-4  
 Growth resistance 439  
 Grnneisen relation 272
- Heat of mixing 178, 180, 187-8, 191, 193, 197-8  
 Helmholtz free energy 129, 200  
 Henry's law 182-3  
 Heterophase fluctuations 428-9, 431  
 Hexagonal close-packed structure 42, 107, 110, 113, 120-4, 127, 298, 308-10, 362, 381  
 Hexagonal lattices 39-42  
 High angle grain boundaries 338, 344-6, 349, 351, 353-5, 360-1, 414, 416  
 Homogeneous transformation 5, 16  
 Hooke's law 72  
 Hume-Rothery rule 172, 207
- Icosahedral packing 162  
 Ideal solution 174, 178-9  
 Image displacements 203-4, 268-70  
 Impingement 17-18, 531-2  
   soft 488, 539-40, 542-3  
 Improper rotation 46, 50  
 Incoherent boundaries 361, 485  
 Incoherent particles 462, 464-6, 469, 476-7  
 Indices  
   Miller-Bravais 39-41, 122  
   of a direction 27  
   of a plane 30  
 Indirect neighbours 121, 163  
 Indium  
   -magnesium alloys 210  
   -silver alloys 210, 304  
 Induction (or incubation) period 16-17, 448  
 Interaction energy  
   of atom pair 113-22, 186-7, 211  
   of dislocation and solute atom 269-72  
     of two dislocations 264-7  
 Interatomic distances in liquids 160-1  
 Interface-controlled growth 7, 480-6  
 Interface mobility 7-8, 484-7, 503-6, 515, 519, 521-3  
 Interlamellar spacing 510-27  
   dependence on supercooling 513  
   factors determining 513-14, 516, 523, 525-6  
   minimum value 511, 513  
 Intermetallic phases 162  
 Interphase boundaries 361-76  
   coherent 361-3, 366  
   diffuse 5, 166-7, 481-2  
   dislocation models 363-76  
   epitaxial 363-4  
   glissile 6-7, 365  
   incoherent 361, 363, 485  
   martensitic 363-4  
   semi-coherent 361-5  
 Interplanar spacing 30, 38, 41  
 Intersection of dislocations 252  
 Interstitial diffusion 378-80, 382, 385-7, 397, 413  
 Interstitial solid solution 171, 269, 271, 378-9  
 Interstitials 127-8, 132-4  
   energy of formation 133-4  
   energy of motion 139  
   equilibrium concentration 130  
   in irradiated crystals 146-7  
 Intrinsic faults 123-4, 126, 296-300, 302, 304-5, 309-10  
 Invariant line 45, 62-3  
   strain 63  
 Invariant plane 45  
   strain 59-63, 291-3, 366  
 Iron 107-11  
   -chromium alloys 110  
   -molybdenum alloys 109  
   -nickel alloys 110  
   -palladium alloys 210  
   -platinum alloys 210  
   -silicon alloys 210  
 Irrational directions 27  
 Ising model 153, 186, 196-7, 216  
 Isokinetic transformations 550-1  
 Isothermal transformation curves 16-21, 532-46  
 Isotope effect in diffusion 92-4, 408-9, 411
- Jogs 238-9, 252-4, 301, 304-5, 307  
 Johnson  
   molecule 393, 404  
   potential 120, 126-7, 131-2, 278, 312, 353

- $K_1, K_2$  planes 52–6  
 Kinetics of transformation  
   diffusion-controlled 538–46  
   grain boundary nucleated 530–8  
   grain corner nucleated 532, 535–8  
   grain edge nucleated 532, 535–8  
   homogeneous 16  
   linear growth laws 16, 529–38  
 Kink density 150–2  
 Kinks in steps 149–52  
 Kirkendall effect 250, 382, 390–1, 394, 403  
 Kopp–Neumann rule 197  
 Kossel  
   crystal 151, 160  
   site 149  
 Kronecker delta 24
- Lambda point 218–19, 221–5  
 Lamellation 509–10  
 Latent heat  
   of melting 155, 160, 164–5, 167, 446  
   of superlattice changes 219, 221–2, 224, 226–8  
   of vaporization 1, 149  
 Lattice 23–31  
   correspondence 57–8, 318–19, 361–4  
   correspondence functions 318–19  
   deformation 56–7, 62–3, 322, 365–76  
   geometry 37–8, 41  
   invariant deformation 322  
   theories of liquids 164–6  
 Lead 76, 135, 413  
 Length of vector 29, 34, 37–8, 41, 47, 65  
 Lennard-Jones forces 117, 199, 354, 423  
 Lenticular plates 14–15  
 Line tension of dislocation 243, 267–8  
 Liquids 160–7  
 Lithium 76, 110, 117, 135  
   –magnesium alloys 210  
 Local Burgers vector 319–21  
 Local vectors 317–18  
 Long period superlattices 227–8  
 Long-range order 162–6, 214–18, 220–1, 223–7  
    $\gamma$ -loop in iron alloys 109  
 Lorentz  
   contraction 282  
   transformation 282
- $M_d$  temperature 14  
 $M_f$  temperature 14  
 $M_s$  temperature 14  
 Magnesium 135, 309, 311  
 Magnetic disordering 108
- Manganese 21, 110  
 Martensitic transformations 13–16, 363–4  
 Matano interface 390–1, 403  
 Matching plane 362  
 Matrix  
   column 24  
   diagonal 44–5  
   eigenvalues of 44  
   orthogonal 46  
   reciprocal 32  
   row 24–5  
   transposed 25  
 Maximum growth rate 499, 501, 504, 513, 516–19, 525  
 Maxwellian atmosphere 272, 507, 545  
 Mechanical stability 115–18  
 Metallic cohesion 111–22  
 Metastable state 4  
 Metric tensor 34–5, 37–8, 41–2  
 Microscopical reversibility, principle of 83, 98, 100  
 Military transformations 6  
 Miller indices 147, 370  
 Miller–Bravais indices 39–41, 122  
 Minimum entropy production, principle of 104–5, 516  
 Mirror planes 51  
 Misfit energy 201  
 Misfitting sphere model 200–8, 269–70, 464, 469  
 Misorientation at grain boundary 328, 338–9, 341, 344  
 Mobility of boundary 7–8, 484–7, 515, 519, 521–3  
   activation enthalpy 485  
   effect of impurities 485  
 Molecularity of reaction 81–2  
 Molybdenum 76, 118  
 Morse potential 118–19, 127, 131–2, 138, 199, 278, 310, 354  
 Motif unit 25–6
- Natural changes 3, 95  
 Nearest neighbour model 111, 116–17, 119–22, 186–8, 208, 211–12, 214, 216, 220–1, 226–8  
 Néel temperature 110  
 Neutron diffraction 161  
 Nickel 76, 131–3, 136, 138  
   –manganese alloys 210  
   –platinum alloys 210  
   –tin alloys 210  
 Non-isothermal reactions 546–52  
 Nucleation  
   at preferred sites 20, 428, 452–61, 472–8  
   Becker–Döring theory 436–41  
   difficulties in classical theory 423, 425, 430–1

- Nucleation (*continued*)  
 effect of strain energy 461–72  
 heterogeneous 428, 452–61  
 homogeneous 428–46  
 in cavities 454–6  
 in condensed assemblies 442–6, 461–78  
 of coherent particles 466–71  
 of dislocations 280–1  
 of incoherent particles 462–6  
 of liquid from vapour 424–41  
 of martensite 470  
 of semi-coherent particles 471–2  
 of solid 441–6  
 of vacancy clusters 142–3, 306–7, 310, 403  
 on dislocations 472–8  
 on grain boundaries 456–61  
 on solid substrate 452–4  
 random 19  
 rate, operational definition 17  
 relative rates on different sites 459–61, 475–6  
 site saturation 533–8  
 steady-state rate 433–5, 439–41, 444–6, 448–52  
 time dependent 19–21, 530, 539  
 transient 446–52  
 two-dimensional 160, 442, 484  
 Volmer–Weber theory 428–36  
 Nucleation and growth transformations 5–6, 11–13
- O*-lattice theory 351, 366 74  
 Onsager reciprocal relations 97–104, 392–3, 407  
 Order  
 long-range 162–6, 214–18, 220–1, 223–7  
 short-range 214–16, 223 5, 228–33  
 Order of reaction 16, 21, 81  
 Order–disorder transformations 208–28  
 analogy with melting 164  
 ground states 211–14  
 superlattice structures 208–28  
 thermodynamic classification of 226–7  
 Ordered solid solutions 171  
 Orthogonal matrix 46  
 Orthohexagonal axes 39
- Pair density (probability) function 172–3, 215, 225, 229  
 Partial excess functions 180–1  
 Partial free energy 175  
 Partition functions 86–9, 91–3, 174, 191–3, 425–7, 430–1  
 Peierls–Nabarro model 273–81, 284  
 Peierls  
 energy 276–8  
 stress 276–80, 308
- Peritectic reaction 178  
 Phases 1–6, 10  
 Phenomenological equations 97  
 of diffusion 392–5, 398–9, 402  
 Physical defects 128  
 Physical property measurements 104, 134–6, 139–40  
 Pile-ups 315  
 Place exchange diffusion 379–80, 387  
 Plastic deformation  
 in diffusion experiments 419  
 in martensitic transformations 14  
 in nucleation and growth transformation 12  
 Platinum 141, 147  
 –silver alloys 208  
 Plutonium 413  
 Point defects 127–47  
 annealing of 139–47  
 condensation of 141–6, 238–9, 300, 305, 307, 310, 403–4  
 generation by dislocations 238–9, 250–4, 403–4  
 mobility of 137–40  
 Point group 173  
 Poisson's distribution 230–1  
 Poisson's ratio 76  
 Polar free energy diagram 155–60, 295  
 Polycrystalline aggregates 76, 329, 333–4  
 Polygonization 403  
 Polymorphic transitions 107–11  
 Porosity 403  
 Potassium 76  
 Precipitation  
 diffusion-controlled 486–508  
 discontinuous (cellular) 508, 519–21, 523, 525–7  
 interface-controlled 486–9, 495, 498  
 on dislocations 506–8, 544–6  
 Preferred orientations 327–8  
 Primitive unit cell 24, 37, 39  
 Principal  
 axes 43, 46–50  
 deformations 50  
 strains 59, 63  
 stresses 71  
 Pseudopotentials 113  
 Pure strain 46–50
- Quasi-chemical theory 191–200  
 Quasi-steady-state distribution 433–7, 446–7  
 Quenching experiments 12, 134–6, 141–6
- Radial distribution function 160–1  
 Radiation damage 146–7  
 Radiation damping 285

- Random close-packing 163
- Random loose-packing 163
- Random walk 140, 383–4, 386, 395–7, 404
- Rate constant 16, 81–3, 86–94, 220, 385, 404–9
- Rayleigh wave 283
- Reaction order 16, 81
- Reaction path 85–7, 89, 92
- Reaction rate theory 85–94, 137, 407–11, 482–3
- Reciprocal
  - lattice 31
  - matrix 32
  - space 31
  - strain ellipsoid 44, 48–9
  - twins 54
  - vectors 27–8, 31, 157
- Reconstructive transformations 6
- Recovery processes 139–47
- Regular solutions 187–91
- Relaxation 128, 131
- Reversibility
  - in martensitic transformations 14
  - in nucleation and growth transformations 12
- Ring diffusion 379–80
- Rodrigues
  - Frank map 46
  - vector 46
- Rolling texture 327
- Rotation 47–8, 50–1, 60, 63
  - improper 46, 50
  - infinitesimal 65, 67
- Saddle point 85, 91–3, 137–9, 464, 466, 471
- Scalar product 28–9, 34–5, 38, 41–2
- Schmid law 256
- Second order phase transitions 161, 214, 226–7
- Second order reaction 81, 141
- Self-diffusion 381, 383–4, 386, 396–7
- Semi-coherent boundaries 361–5
  - epitaxial 363–4
  - glissile 6–7, 365
  - martensitic 363–4
- Semi-coherent particles 471–2
- Sessile dislocations 288, 296–7, 300–2, 309–10
- Shape
  - change in transformations 12–15
  - deformation 321–2
  - equilibrium 155–60
  - factor 424, 443
  - of new phase 12–15, 61
  - perturbation 495–8
- Shear
  - pure 53–4
  - simple 52–4, 60
  - strain 65, 72
  - strain (engineering) 72
- Shear modulus 75, 129, 165, 203, 206
  - temperature variation 206, 410
- Shockley partial dislocation 296, 300, 302, 305–6, 308, 310
- Short-range order 214–16, 223–5, 228–33
  - coefficients 223, 225
- Shuffles 52
- Silicon 76, 250, 313, 413
- Silver 76, 113, 135, 299, 416, 419
  - zinc alloys 210, 227
- Similarity transform 43
- Singular surfaces 158–9, 480–1
- Sinks for point defects 139–45, 403
- Size factor 172, 207
- Sodium 76, 117, 119, 135
- Solid–liquid interface 166–7
- Solid solution 26, 271–2
- Solubility gap 190
- Solubility limits 175, 180–2, 189–90, 194–5
- Somigliana dislocations 257–8
- Space group 26
- Specific heats 107–10, 219–20, 223–4, 226–7
- Spinodal 184–5, 190, 207, 398–400
- Stability of growth front 495–8, 525–6
- Stabilization 15
- Stacking faults 122–7, 296–312
  - energy of 125–7, 298–300
  - tetrahedron 306–7, 315
- Stacking operators 123
- Stair-rod dislocation 301–3, 305–7
- Steady-state phenomena 95–6, 104
- Step growth 480–2, 484, 505–6
- Steps
  - on crystal surface 147–52
  - on interface 288–94, 505–6
- Stiffness constants 73–6
- Strain ellipsoid 43–4, 48–50
- Strain energy 73–4
  - accommodation factor 464, 469–71
  - dependence on particle shape 465–6, 468–72
  - function in elastic theory 73–4
  - in solid solutions 200–7
  - misfitting sphere model 269–70, 464, 469
  - of coherent particles 466–71
  - of dislocations 261–7, 275, 294
  - of incoherent particles 464–6
  - of martensite 470
  - using anisotropic elastic theory 262, 267, 466, 470
- Strain quadric 66
- Strain tensor 64, 67–8
- Stress quadric 72
- Stress tensor 69–70

- Stress vector 69–70
- Structure sensitive properties 3
- Sub-regular solutions 207–8
- Substitutional solid solution 26, 171, 378–9
- Substructure 314–16
- Supercooling 445–6, 502
- Superheating 445
- Superlattice structures 208–28
- Superposition, principle of 64–5
- Supersaturation ratio 427, 433, 435
- Surface dislocation 322, 365–6, 376
- Surface forces 69
- Surface free energy 4–5, 328–9, 422–8
  - deduced from nucleation experiments 423
  - of coherent interface 359–60, 462, 471
  - of curved interface 328–9, 423
  - of grain boundary 164–5, 328–9, 356–61
  - of liquid–solid interface 164–6
  - of stacking fault 126–7, 298–300, 304–11
  - of stepped surface 147–8
  - polar diagram of 155–60
- Surface melting 154
- Surface node 376
- Surface roughness 152–3
- Surface steps 147–52
- Surface stress 329
- Surface tension 329
  
- Tensor notation 35–8
- Tensors
  - antisymmetric 65
  - contravariant 35–6
  - covariant 35–6
  - fourth rank 72
  - matrix notation 42–3
  - metric 37–8
  - mixed 36, 43
  - symmetric 46
  - transformation of axes 35–7, 43–5
- Texture 327
- Thallium 381–2, 413
- Thermodynamic
  - classification of transformations 226–7
  - currents (or fluxes) 96–8, 101–4, 484
  - equations of motion 97
  - forces 96–9, 101–4, 484
- Thermodynamics of irreversible processes 95–105, 483–4, 516
- Thermoelastic damping 285
- Thompson tetrahedron 303–4, 306–7
- Thomson
  - Freundlich equation 184, 427
  - Helmholtz equation 184
- Thorium 76
- Tilt boundaries 335–9, 351–2, 354–9
  - asymmetrical 338–9, 354, 359
  - dislocation model 336–9
  - energy of 356–9
  - symmetrical 336, 338, 345, 348, 351, 354
- Time–temperature–transformation (*TTT*)
  - diagram 546–9
- Time to half-transformation 536–8
- Tin 413
- Titanium 413
- Topological principles 330, 333–4
- Torsion tensor 320–1
- Tracer diffusion 378, 383–4, 386, 394–7
- Traction 69
- Transformation of axes
  - of base vectors 31–2
  - of components of plane normals 33–4
  - of tensor components 35–7, 43–5
  - of vector components 32, 35–7
- Transformations
  - classification 1–16
  - general definition 3
  - heterogeneous 4–6
  - isothermal kinetics 16–21, 529–46
  - non-isothermal kinetics 546–52
  - of thin sheets and wires 20, 546
- Transition state theory 85–94
- Transmission probability 87–8
- Transport processes 96–7, 104
- Trapping of defects 138, 146–7
- Tungsten (wolfram) 76, 118
- Tunnel effect 88
- Twin boundaries 353–4, 359, 366
- Twinning 50–6, 288–94
  - axis 51
  - dislocations 288–94
  - glide 52–6
  - in b.c.c. metals 353–4
  - plane 51
  - shear 52–6
  - types I and II 54–5
- Twins
  - compound 55
  - conjugate 54
  - deformation 52
  - reciprocal 54
  - reflection 51
  - rotation 51
- Twist boundary 340–1
  
- Unit cell 24–7, 37–8, 171–2, 209–10
- Unit step in reaction 81–3, 85–6

- Unrotated planes 45, 53-4, 59, 61-2  
 Unrotated vectors 44-6  
 Uphill diffusion 384, 397-400  
 Uranium 413
- Vacancies 127-32  
   annealing of 139-47  
   energy of formation 128-32, 135-6, 143-6, 410-13  
   energy of motion 137-9, 142-4, 409-12  
   entropy of formation 129-30, 136, 411, 413  
   entropy of motion 129, 137-8, 409-11  
   equilibrium concentration 130, 134-7  
   excess concentration 139-40, 307, 310, 403, 418-19
- Vacancy  
   clusters 131-2, 141-6, 307, 403  
   condensation 141-6, 300, 305, 307, 310, 403-4  
   diffusion 379-83, 385-6, 392-7, 400-7, 409-13
- Van der Waals forces 111, 114, 117
- Vanadium 413
- Variational principles 104, 516
- Vector  
   axial 65  
   base 23-4, 31, 35-41  
   Burgers 236, 238, 240-6, 248-50, 252-3, 255, 258, 260, 262-6, 269, 274-5, 286-301, 303-4, 307-11, 313, 316-21, 336-7, 339, 341-5, 349, 351-2, 360-1, 363-7, 370, 372-6, 476-7  
   column 24  
   components of 24, 35-6  
   contravariant 28, 35-6  
   displacement 63-8  
   lattice 23-4  
   length of 29, 34, 37-8, 41, 47, 65  
   local 317-18  
   position 63  
   primitive 24, 55  
   reciprocal 27-8, 31, 157  
   row 25  
   scalar product 28-9, 34-5, 38, 41-2  
   scalar triple product 27  
   stress 69-70  
   unrotated 44-6  
   vector product 29, 38, 42
- Vegard's law 172, 205
- Velocity of dislocation 282-6
- Velocity of sound 282-4, 286
- Virtual glide surface 247, 251
- Volterra dislocations 257-8, 260
- Volume change during melting 160
- Volume ratio 47-8, 53, 61, 67
- Voronoi polygon 120
- Water 441
- Wave vector of fluctuation 399-400, 497
- Widmanstätten precipitation 13, 490, 498
- Wigner-Seitz  
   cell 356  
   method 543
- Wigner-Seitz-Slater  
   cell 372  
   method 120, 122
- Wolfram *see* Tungsten
- Wulff's theorem 155-60, 295, 441
- X-ray diffraction 23, 107, 135, 171-3, 205, 207, 225, 228-9, 313-15  
   from liquids 160-1
- Young's modulus 76
- Zener-Hillert growth model 498-505
- Zener theory of diffusion 89-90, 408-10
- Zero-point energy 106
- Zinc 309, 376, 381-2

## *Note on Symbols used for Physical Quantities*

Reprinted from the Second Edition, reprinted with corrections (of Part I)

THE number of physical quantities introduced in the text is so large that a separate symbol for each would have required many different type faces. Some duplication has therefore been accepted, but this has been minimized by suitable use of superscripts and subscripts, and it is believed that no confusion can arise. Wherever possible, the recommendations of the appropriate international commissions on nomenclature have been followed, but there are inevitably some variations. The main features of the notation are given in the following list, and a complete index of symbols is given at the end of each volume.

<i>Type of quantity</i>	<i>Type face</i>	<i>Example</i>
Universal physical constants	Bold face italic	$\mathbf{R}, \mathbf{k}, N_0$
Scalar quantities	Italic, Roman or Greek	$x, T, \alpha$
Vectors or higher order tensors	Bold face upright, Roman or Greek	$\mathbf{u}, \mathbf{S}, \xi$
Vector or tensor components	Italic, Roman or Greek	$u_i, S_{ij}, \xi_i$
Matrix representations of vectors or tensors in particular bases	Sans serif, Roman <sup>†</sup>	$\mathbf{u} = [\Lambda; \mathbf{u}]$ $\mathbf{S} = (\Lambda \mathbf{S} \Lambda)$
Sets of base vectors	Small capitals	$\Lambda = \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$

The meanings of the main superscript and subscript symbols are given in the following list.

<i>Superscript before main symbol</i>	<i>Meaning</i>
$v$	per unit volume
$L$	per unit length
$B$	per unit area of boundary
$S$	per unit area of substrate
$E$	per unit length of edge
$C$	per grain corner

<sup>†</sup>The bold face symbol  $\mathbf{I}$  is used for both the vector  $\mathbf{I}$  and for its matrix representation. This avoids confusion with the sans serif symbol  $\mathbf{I}$  which represents the unit matrix. Matrix representations of vectors denoted by Greek letters are avoided if possible; when they occur, the bold face symbol is used.

<i>Superscript after main symbol</i>	<i>Meaning</i>
$v, l, s, \alpha, \beta, \gamma, \text{ etc.}$	relating to vapour, liquid, solid, $\alpha, \beta, \gamma, \text{ etc.}$ phase
$o$	pure substance
$H$	homogeneous
$B, E, C, S$	relating to a grain boundary, edge or corner or to a surface
$\theta$	maximum value
$m$	matrix or metastable phase
$\alpha\beta$	relating to an $\alpha$ - $\beta$ interface
$\alpha\beta$ (thermodynamic quantities)	from $\alpha$ to $\beta$ . Thus $\Delta g^{\alpha\beta} = g^{\beta} - g^{\alpha}$
<i>Subscript after main symbol</i>	
$i, j, k$	identifying subscript for vector components
$ij, kl, \text{ etc.}$	identifying subscript for tensor components, or for transformation matrices
$A, B, C, \dots$	components of the assembly (elements)
$T$	at temperature $T$
$\square, \square\square, \bullet$	vacancy, divacancy, interstitial
$c$	critical value of a quantity
<i>Subscript before main symbol</i> (used only after mathematical symbol $\Delta$ )	
$a$	activation, e.g. $\Delta_a g =$ free energy of activation
$m, n$	mixing, e.g. $\Delta_m g =$ free energy of mixing
$e, f$	excess, e.g. $\Delta_e s =$ excess entropy of mixing

The summation convention is used with the subscript notation for vector and tensor components. Contravariant and covariant components are not distinguished by the positions of the sub- or superscript, but components referred to a reciprocal set of base vectors (covariant components) are distinguished by an asterisk,  $u_j^*$ . The asterisk is omitted for the components of plane normals, which are always referred to the reciprocal basis. Column and row vectors are distinguished by square and round brackets respectively. The transposed matrix of  $\mathbf{A}$  is written  $\mathbf{A}'$ , and the inverse matrix  $\mathbf{A}^{-1}$ . These features of the tensor and matrix notation are explained more fully in the text.

## *Index of Symbols*

Reprinted from the Second Edition, reprinted with corrections (of Part I)

<i>Symbol</i>	<i>Meaning</i>
$a, a_0$	Edge length of unit cell
$a$	Coefficients of pairwise potential
$a^B, a^E, a^C$	Kinetic parameters
$a$	Kinetic parameter
$\mathbf{a}_i$	Base vectors
$\mathbf{a}_i^*$	Reciprocal base vectors
$\mathbf{a}^+$	Reciprocal (four-axis) base vectors
$\bar{\mathbf{a}}_i$	Eigenvectors
$\mathbf{A}$	Component of alloy
$\mathbf{A}^+$	Activated complex
$A$	Force constant
$A_1, A_2$	Coefficients in equation for $A_m g$
$A$	Constant in equation for grain boundary energy
$A(b)$	Coefficients in solution of diffusion equation
$A_n$	Coefficients in series solution of diffusion equation
$\Lambda$	Base vectors $\mathbf{a}_i$ , defining a co-ordinate system
$A_{ij}$	Dislocation tensor
$A(\beta, t)$	Amplitude of component of composition fluctuation
$\mathbf{A}$	Two-dimensional rotation matrix
$b$	Number of vacancy-vacancy bonds in cluster
$b$	Continuous variable in solution of diffusion equation
$b_n$	Variable in series solution of diffusion equation
$b^B, b^E, b^C$	Kinetic parameters
$\mathbf{b}_i$	Base vectors
$\mathbf{b}$	Burgers vector
$\mathbf{b}_e, \mathbf{b}_s$	Edge and screw components of $\mathbf{b}$
$\mathbf{b}^{(1)}$	Burgers vector of primary interface dislocation
$\mathbf{b}^{(2)}$	Burgers vector of secondary interface dislocation
$B$	Force constant
$B_e, B_s$	Constants in elastic theory of dislocations
$B_{ij}$	Surface dislocation tensor
$B$	Factor determining $D_{A^-}/D_{A^+}$

<i>Symbol</i>	<i>Meaning</i>
<b>B</b>	Relation between two O-lattices
$c$	Velocity of sound
$c$	Edge length of unit cell
$c_A, c_B$	Concentration of components $A, B$
$c^+$	Concentration of activated complexes
$c_0$	Average solute concentration
$\bar{c}$	Mean solute concentration over a plane
$c_0$	Initial solute concentration
$c$	Total solute concentration
$c_{A1}$	Concentration of $A$ on plane 1
$\bar{c}_{\square}$	Equilibrium vacancy concentration
$c^{\alpha}, c^{\beta}$	Concentration of solute in $\alpha$ and $\beta$ phases in contact at an interface
$c^m$	Initial solute concentration in metastable phase
$c^{\infty}(t)$	Time dependent concentration of solute in regions remote from interface
$\Delta c$	Effective concentration difference producing diffusion
<b>c</b>	Base vector
$c, c_{ijkl}$	Elastic stiffnesses (or elastic constants) tensor
$c, c_{ij}$	Elastic stiffnesses ( $i, j = 1 \dots 6$ )
$C_p$	Specific heat at constant pressure
$\Delta C_p^{\alpha\beta}$	Difference in specific heats = $C_p^{\beta} - C_p^{\alpha}$
$C_1 - C_{37}$	Arbitrary or temporarily unknown constants introduced at various places in text. Only the more important are listed separately below
$C_6, C'_6$	Constants used in elastic theory of misfitting particle
$C_7$	Linear temperature dependence of enthalpy of diffusion
$C_8$	Constant in distribution function for embryos
$C_9$	Constant in equation for nucleation rate
$C_{14}$	Constant in equation for diffusion controlled growth
$C_{16}$	Constant in equation for solute-dislocation interaction
$C_{17}$	Constant in equation for cellular growth
$C_{20}$	Constant in equation for time-dependent nucleation
$C, (\Lambda \text{ C } B)$	Correspondence matrix relating vectors of two lattices with bases $\Lambda$ and $B$
$d_h, d$	Interplanar spacing of lattice planes
$d$	Height of step in interface or surface
$d$	Distance moved after successful activation
<b>dd</b>	Line element of dislocation
$D$	Force constant
$D, \mathbf{D}$	Diffusion coefficient
$D_{\text{chem}}$	Chemical diffusion coefficient
$D_A, D_B$	Diffusion coefficients for components
$D_{A+}$	Self diffusion coefficient of $A$
$D_{A.A}, \text{etc.}$	Coefficients in theory of diffusion
$D_{\square}$	Diffusion coefficient for vacancies
$D_0$	Pre-exponential term in equation for $D$
$D_{01}, D_{02}$	Pre-exponential factors from mono- and divacancies

<i>Symbol</i>	<i>Meaning</i>
$\bar{D}$	Mean diffusion coefficient
$D^B$	Grain boundary diffusion coefficient
$D^P$	Diffusion coefficient along dislocations
$D_n$	Parameter in theory of nucleation
$\mathbf{D}, D_{ij}$	Generating deformation relating local lattice to reference lattice
$e_i$	Fluctuations in an assembly at equilibrium
$e_\mu$	Fraction change of $e_\mu$ with atom fraction of solute
$\mathbf{e}$	Displacement vector in invariant plane strain, and especially in shape deformation
$\mathbf{e}, e_{ij}$	Strain tensor
$e_{ij}^I$	Stress free strains in unconstrained transformation
$e_{ij}^C$	Matrix strains in constrained transformation
$e_{ij}$	Inclusion strains in constrained transformation
$E$	Experimental activation energy
$E_n$	An embryo of $n$ atoms
$f_f, f_b$	Functions in rate equation
$f_{AA}, f_{BB}$	Constants in theory of conformal solutions
$f(\theta)$	Polar free energy curve
$f^B, f^E, f^C$	Functions in formal theory of transformation kinetics
$f(t)$	Isothermal transformation function
$f, \mathbf{f}$	Force per unit length of dislocation
$f_g$	Glide component of force on dislocation
$f_r, f_0$	Components of force between two dislocations
$f$	Correlation factor
$f(\bar{p})$	Function used in Horvay-Cahn growth theory
$\mathbf{f}$	Fault displacement vector
$F$	Function of jump probabilities
$F$	Helmholtz free energy
$F$	Function in theory of non-isothermal kinetics
$g_{ij}$	A positive definite form
$g_{AA}, g_{BB}, g_{AB}$	Constants in theory of conformal solutions
$g(x_1)$	Function used in solution of diffusion equation
$g(\zeta)$	Function of volume fraction transformed
$g^l, g^s, g^\alpha$ etc.	Gibbs function per atom in liquid, solid, $\alpha$ phase, etc.
$g_{A^\alpha}$	Gibbs function per atom of pure $A$ in $\alpha$ phase
$g_A^\alpha$	Chemical potential per atom of $A$
$g_{Ar}^\alpha, g_{Ar}$	Effective chemical potentials for $\beta$ particles of radius $r$ in $\alpha$ matrix
$g_{ih}$	Free energy per atom of inhomogeneous solid solution
$\Delta_m g^\alpha$	Free energy of mixing per atom in $\alpha$ phase
$\Delta_n g^\alpha$	Free energy of mixing per atom in $\alpha$ phase, defined relative to $A$ in $\alpha$ and $B$ in $\beta$ phase
$\Delta_e g$	Excess free energy of mixing
$\Delta_f g$	Excess free energy of mixing with same zero as $\Delta_n g$
$\Delta g_T^{\alpha\beta}$	$g^\beta - g^\alpha$ at temperature $T$ for pure component
$\Delta g_s$	Elastic strain energy per atom of transformed region
$\Delta g_{\square}, \Delta g_{\square\square}$	Free energy of formation of vacancy, divacancy

<i>Symbol</i>	<i>Meaning</i>
$\Delta g_{\bullet}$	Free energy of formation of interstitial
$\Delta g^{\ddagger}$	Free energy of activation per atom
$\Delta_a g$	Free energy of activation per atom
$\Delta_a g^{\ddagger}$	Free energy of activation, especially for coherent growth
$\Delta_a g^*$	Free energy of activation, especially for incoherent growth
$\Delta_a g_{\square}$	Free energy of activation for vacancy migration
$\Delta g^{\Gamma}$	Change in free energy per atom produced by a transformation
$\Delta g^{\gamma}$	Change in free energy per atom produced by motion of an interface
$\mathbf{g}$	Body force vector
$G$	Gibbs free energy
$\Delta G, \Delta G_n$	Free energy change to form embryo of $n$ atoms
$\Delta G'$	Chemical free energy of embryo formation
$\Delta G_{tr}, \Delta G_{rot}, \Delta G_{rep}$	Additional contributions to free energy of embryo formation
$\Delta G_c, \Delta G_c^H$	Critical free energy for (homogeneous) nucleation
$\Delta G_c^B$	Critical free energy for boundary nucleation
$\Delta G_c^E, \Delta G_c^C$	Critical free energy for nucleation on grain edges, corners
$\Delta G_c^D$	Critical free energy for nucleation on a dislocation
$\Delta G^{\ddagger}$	Molar free energy of activation
$\Delta_a G$	Molar free energy of activation
$G$	Function in theory of non-isothermal kinetics
$\mathbf{G}, G_{ij}$	Metric tensor
$h$	Planck's constant
$h$	Height of a cavity
$h(x_2)$	Function used in solution of diffusion equation
$h(T)$	Function of temperature
$\Delta h_0^{\alpha\beta}$	Difference in enthalpies per atom of $\beta$ and $\alpha$ phases at 0 K
$\Delta h^{vv}$	Heat of vaporization per atom
$\Delta h^{sl}$	Latent heat of melting per atom
$\Delta_m h$	Heat of mixing per atom
$\Delta_m h_A$	Partial heat of mixing per $A$ atom
$\Delta h_{\square}, \Delta h_{\square\square}$	Enthalpy of formation of vacancy, divacancy
$\Delta h_n$	Formation energy of vacancy cluster
$\Delta h_1, \Delta h_2$	Contributions to enthalpy of diffusion from vacancies and divacancies
$\Delta h_{\bullet}$	Enthalpy of formation of interstitial
$\Delta_a h$	Enthalpy of activation
$\Delta_a h^*$	Enthalpy of activation, especially for incoherent growth
$\mathbf{h}$	Normal to a set of lattice planes
$h_i, {}^A h_i$	Components of $\mathbf{h}$ referred to reciprocal base $\Lambda^*$
$\mathbf{h}', (\mathbf{h}; \Lambda^*)$	Representation of $\mathbf{h}$ as a row matrix
$H$	Enthalpy function
$\Delta_m H$	Heat of mixing
$\Delta H^{\ddagger}$	Molar enthalpy of activation
$\Delta_a H$	Molar enthalpy of activation
$H$	Function in theory of non-isothermal kinetics
$i$	Molecularity of unit step

<i>Symbol</i>	<i>Meaning</i>
$i$	Supersaturation ratio
$i$	Dimensionality of nucleation site
$\mathbf{i}_i$	Set of orthonormal vectors
$\mathbf{i}$	Direction of dislocation line element
$I, \mathbf{I}$	Generalized current or flux
$I_A$	Flux of $A$ atoms
$I_{\square}$	Flux of vacancies
$I$	Steady state nucleation rate
$I(t)$	Time-dependent nucleation rate
${}^v I, {}^o I^H$	Homogeneous nucleation rate per unit volume
${}^v I^B, {}^v I^E, {}^v I^C$	Grain boundary, edge and corner nucleation rate per unit volume of assembly
$S I, {}^B I$	Nucleation rate per unit area of substrate, grain boundary
${}^E I$	Nucleation rate per unit length of grain edge
${}^C I$	Nucleation rate per grain corner
${}^v I^D$	Dislocation nucleation rate per unit volume of assembly
$I$	Moment of inertia
$I, I'$	Cell factors for centred unit cells
$\mathbf{i}$	Orthonormal basis $\mathbf{i}_i$ of co-ordinate system
$\mathbf{1}, \delta_{ij}$	Unit tensor
$I_i$	Flux of $i$ th component
$j$	Degree of supersaturation
$\mathbf{J}, (\mathbf{A} \mathbf{J} \mathbf{B})$	Transformation matrix relating bases $A$ and $B$
$J_{ij}$	Components of $\mathbf{J}$
$k$	Boltzmann's constant
$k, k_f, k_b$	Reaction rate constants
$k_A$	Frequency of motion of $A$ atom into adjacent vacant site
$k_{A1}, k_{A2}$	Jump frequencies of specified atoms into vacant site
$k_c$	Jump frequency for dissociation of vacancy $\mathbf{B}$ atom complex
$k_r, k_w$	Rate constants in theory of ordering
$k$	Constant in Avrami equation
$k^S$	Interface step density
$\mathbf{k}^F$	Wave vector at Fermi surface
$K, K_c^*, K_c$	Equilibrium constants
$K$	Bulk modulus of elasticity
$K_B$	Bulk modulus of component $B$
$K_1$	Twinning (invariant) plane of a mechanical twin
$K_2$	Conjugate twinning plane (second undistorted plane) of a mechanical twin
$K$	Gradient energy coefficient
$K$	Factor in anisotropic elastic energy of dislocation
$K^B$	Interface concentration ratio in cellular growth
$K_0$	Solute-boundary interaction coefficient
$\Delta K$	Parameter in diffusion theory
$l$	Edge length of nucleus
$e$	Curvilinear coordinate along interface

<i>Symbol</i>	<i>Meaning</i>
$\mathbf{l}$	Local Burgers vector
$\mathbf{l}_+, \mathbf{l}_-$	Axes of rotation
$L$	Separation of elements of disclination dipole
$L_1, L_2$	Macroscopic lengths
$L$	Long range order parameter
$L^B$	Mean grain diameter
${}^v L^E$	Grain edge length per unit volume
${}^v L^D$	Length of dislocation line per unit volume (dislocation density)
$L^\beta$	Length of a line intersected by $\beta$ regions
$L_c^\beta$	Extended line intercept
$L^P$	Value of $l$ at mid-points of plates
$\mathbf{L}, (\mathbf{L} \ \kappa)$	Orthogonal matrix relating bases which are connected by a pure rotation
$L_{ij}$	Direction cosines (elements of $\mathbf{L}$ )
$L_{klm}$	Coefficients of connection
$m_i$	Atomic mass of $i$ th atom
$m$	Effective mass of complex
$m$	Attractive exponent in force law
$m$	Arbitrary constant
$\mathbf{m}_i$	Vector reciprocal to unit normal $\mathbf{n}_i$
$M^\ddagger$	Activated complex
$M$	Mobility or conductance
$\mathbf{M}, M_{ij}$	Generalized mobility or conductance
$M'_{ij}$	Atomic mobilities
$M''_{ij}$	Diffusional mobilities per unit volume
$M^B$	Boundary mobility
$M_s$	Martensite start temperature
$M_d$	Temperature below which martensite forms under stress
$M$	Number of atoms per unit area of planar diffusion source
$M^c$	Boundary mobility per unit concentration difference
$n$	Any undetermined small number
$n$	Number of vacancies
$n$	Number of atoms in embryo or nucleus
$n_c$	Number of atoms in nucleus of critical size
$n$	Repulsive exponent in force law
$n$	Time exponent in Avrami equation
$n_{ik}$	Number of paths with $(i-1)$ jumps to site $k$
$\mathbf{n}$	Unit vector normal to surface or lattice plane
$N_0$	Avogadro's number
$N$	Any undetermined large number
$N$	Total number of atoms (molecules) in assembly
$N_A$	Total number of $A$ atoms in assembly
$N^v, N^\alpha$	Total number of atoms in vapour phase, $\alpha$ phase, etc.
$N^S$	Number of atoms adsorbed on substrate
${}^v N^S$	Number of atoms on substrate per unit volume of assembly
${}^v N^D$	Number of atoms at dislocation cores per unit volume of assembly

<i>Symbol</i>	<i>Meaning</i>
${}^v N^C$	Number of grain corners per unit volume
${}^v N_0$	Number of pre-existing nuclei or nucleation sites
$N_n$	Number of embryos of size $n$ in stable assembly
$N_{AB}$	Number of unlike nearest neighbour contacts per unit coordination number
$N_{AA}, N_{BB}$	Number of like nearest neighbour contacts per unit coordination number
$N_{AB}$	Equilibrium value of $N_{AB}$
$N_{AB}^0$	Maximum value of $N_{AB}$
$o_n$	Number of surface atoms in embryo $E_n$
$O$	Area of a face
$O^b$	Area of a boundary
$O^B$	Total area of grain boundaries
${}^v O^B$	Grain boundary area per unit volume of assembly
$O^S$	Area of substrate
${}^v O^S$	Substrate area per unit volume of assembly
$O^\beta$	Area of plane intersected by $\beta$ regions
$O_c^B$	Extended area of plane
$O_n$	Surface area of an embryo $E_n$
$O_c$	Surface area of nucleus of critical size
$p$	Pressure
$p_r, p_\infty$	Vapour pressure in equilibrium with surfaces of radii $r, \dots, \infty$
$p$	Elective pressure
$\bar{p}$	Péclet number for diffusional growth
$\bar{p}^S$	Péclet number for step growth
$p$	Minimum number of atoms in heterophase fluctuation
$\mathbf{p}$	Stress vector
$\mathbf{P}$	A large vector in a boundary
$P$	Probability of a fluctuation
$P_{AA}(r), P_{AB}(r)$	Pair density functions
$P$	Probability of a site being vacant
$P'$	Fraction of free energy released in a transformation
$P$	Measured value of some physical property, especially electrical resistance
$P_k$	Jump probability for direction $\theta_{1(k)}$
$\mathbf{P}$	Pure deformation tensor, especially pure lattice deformation
$\bar{\mathbf{p}}$	Representation of $\mathbf{P}$ in principal axis system
$dq$	Heat flow
$q_0$	Probability per unit area per unit time of an atom adding itself to an embryo
$q_n$	Probability per unit area per unit time of an atom leaving an embryo of $n$ atoms
$q$	Dimensionless parameter relating interface and diffusion control
$q^S$	Corresponding dimensionless parameter for step growth
$q_i$	Normal coordinates
$Q'_A, Q'_B$	Partition functions referred to a common energy zero

<i>Symbol</i>	<i>Meaning</i>
$Q_A, Q_B$	Partition functions referred to separate zeros
$Q^\ddagger, Q^{\ddagger'}, Q^{\ddagger''}$	Partition functions of activated complex
$Q_x$	Translational partition function
$Q_v$	Vibrational partition function
$Q_{tr}$	Translational partition function for embryos
$Q_{rot}$	Rotational partition function for embryos
$Q_{rep}$	Replacement partition function for embryos
$Q_n$	Partition function for embryo in vapour phase
$\underline{Q}'$	Fraction of excess solute precipitated
$\overline{Q}$	Modified value of $Q'$
$r$	Polar co-ordinate
$r_0$	Equilibrium separation of two atoms
$r_1$	Nearest neighbour separation in crystal
$r_2, r_3$	Second and third nearest neighbour separations
$r_A^0$	Atomic radius in pure $A$
$r_B$	Radius of $B$ atom in infinite crystal of $A$
$r_{AA}^*, r_{AB}^*$	Characteristic atomic lengths
$r_0$	Limiting radius for Hookean displacements around a centre of dilatation
$r_i$	Radius of dislocation core
$r_c$	Radius of dislocation strain field
$r$	Radius of droplet or spherical nucleus
$r_c$	Radius of critical nucleus
$r$	Co-ordinate normal to moving interface
$r^I$	Value of $r$ giving position of interface
$r$	Final radius of transformed region
$r_A$	Probability of an $A$ atom being in a "right" site of a superlattice
$r$	Critical radius for given supersaturation
$r$	Critical semi-thickness of plate or needle
$\mathbf{r}$	Position vector
$R$	Gas constant
$R$	Radius of sphere
$R$	Semi-axis of ellipsoid representing particles of varying shape
$R_n$	Growth resistance
$R^B$	Parameter depending on temperature and grain size
$R$	Radius of flow line
$R'$	Fraction of chemical driving force converted to surface energy
$R(\beta)$	Amplification factor
$R_D, R^P$	Parameters in theory of cellular growth
$\mathbf{R}$	Pure rotation tensor
$\mathbf{R}_+, \mathbf{R}_-$	Orthogonal matrices
$s$	Amount of shear, especially twinning shear
$s$	Unpaired spin per atom
$s$	Limiting nucleus size above which $Z_n = 0$
$\Delta_m^s$	Entropy of mixing per atom
$\Delta_c^s$	Excess entropy of mixing per atom

<i>Symbol</i>	<i>Meaning</i>
$\Delta_c s_A$	Excess partial entropy of mixing per atom of $A$
$\Delta s_{\square}, \Delta s_{\square\square}$	Entropy of formation of a vacancy, divancy
$\Delta s_1, \Delta s_2$	Contributions to entropy of diffusion from vacancies and divacancies
$\Delta s_{\bullet}$	Entropy of formation of an interstitial
$\Delta_a s$	Entropy of activation per atom
$s, s_{ijkl}$	Elastic compliances (or elastic moduli) tensor
$s, s_{ij}$	Elastic compliances ( $i, j = 1 \dots 6$ )
$s'_{ijmn}$	Coefficients in elastic theory of constrained transformation
$\Delta S^{\ddagger}$	Molar entropy of activation
$\Delta_a S$	Molar entropy of activation
$d_c S$	Change in entropy of environment
$d_i S$	Irreversible entropy created during a natural process
$S_1(p), S_2(p)$	Functions used in Trivedi theory of growth
$\mathbf{S}$	Lattice deformation tensor
$\Lambda \mathbf{S} = (\Lambda \mathbf{S} \Lambda)$	Matrix representation of $\mathbf{S}$ in basis $\Lambda$
$\bar{\mathbf{S}}$	Diagonal matrix representation of $\mathbf{S}$ in principal axis system
$t$	Time
$t_{\zeta}$	Time to fixed volume fraction $\zeta$ of transformation
$t_a(T)$	Time to volume fraction $\zeta_a$ of transformation at constant temperature $T$
$\mathbf{t}$	Translation vector in lattice relations
$T$	Temperature
$T$	Line tension of dislocation
$T_c$	Curie temperature
$T_c$	Critical temperature for surface disordering
$T_{\lambda}$	Disordering temperature of a superlattice
$T^s$	Melting point (solidus) temperature
$T^l$	Freezing point (liquidus) temperature
$\Delta T^-, \Delta T^+$	Supercooling or superheating from equilibrium temperature
$T_{ijk}$	Torsion tensor
$\Delta_m u$	Internal energy of mixing per atom
$u$	Velocity of a dislocation
$u_0$	Parameter in empirical equation for dislocation velocity
$u$	Mean velocity of step
$\mathbf{u}$	Lattice vector
$u_i, u_i^*$	Contravariant and covariant components of $\mathbf{u}$
$\mathbf{u}^{(0)}$	Vector of DSC lattice
$U$	Internal energy
$U_0, U_{\infty}$	Internal energy at absolute zero, and at very high temperatures
$U_T, U_{\lambda}$	Internal energy at temperature $T, T_{\lambda}$
$\Delta_m U$	Energy of mixing
$\mathbf{u}, [\Lambda; \mathbf{u}]$	Column matrix of components of $\mathbf{u}$ in base $\Lambda$
$\mathbf{u}', (\mathbf{u}; \Lambda)$	Row matrix of components of $\mathbf{u}$ in base $\Lambda$
$\Delta U^{\ddagger}$	Molar energy of activation
$U^B$	Parameter in theory of cellular growth

<i>Symbol</i>	<i>Meaning</i>
$U_D$	Deformation energy in pairwise model
$U_i$	Orthogonal matrix representing symmetry operation
$U$	Unimodular matrix
$v_a, v_a^*$	Volume of unit cell and reciprocal cell
$v$	Atomic volume
$v_A^0$	Atomic volume of pure $A$
$v_B$	Volume of $B$ dissolved in $A$
$v^l, v^\alpha$ , etc.	Volume per atom in liquid, alpha phase, etc.
$\Delta v^\infty$	Change in volume within closed surface around solute $B$ atom
$\Delta v_{AB}$	Difference in atomic volumes of pure $B$ and pure $A$
$\Delta v^i$	Increase of volume due to image forces in elastic model
$\Delta v$	Total difference in volume of assembly in elastic model
$v_{AB}$	Mean volume per atom in solid solution
$v_\tau$	Volume at time $t$ of transformed region which nucleated at time $\tau$
$v_d$	A characteristic volume in bulk liquid
$\mathbf{v}$	Lattice vector, especially vector formed from $\mathbf{u}$ by a lattice deformation
$v_i, v_i^*$	Contravariant and covariant components of $\mathbf{v}$
$V$	Volume of whole assembly
$V^\beta(t)$	Volume of $\beta$ phase at time $t$
$w_A, w$	Probability of an $A$ atom being in a wrong superlattice site
$\mathbf{w}$	Displacement vector
$W$	Potential energy per atom in solid
$W_s$	Total strain energy in misfitting sphere model
$W_B$	Strain energy in compressed $B$ sphere
$W_A$	Strain energy in surrounding $A$ matrix
${}^l W_d$	Dislocation energy per unit length
${}^l W_s$	Dislocation elastic energy per unit length
${}^l W_c$	Dislocation core energy per unit length
${}^l W_i$	Interaction energy per unit length of dislocation
$W_i$	Interaction potential between dislocation and solute
$W_d$	Energy of dislocation loop
$W_c$	Critical activation energy for formation of a loop under stress
${}^l W'_S$	Strain energy per unit length of moving dislocation
${}^l W'_k$	Kinetic energy per unit length of moving dislocation
${}^l W'_d$	Total energy per unit length of moving dislocation
${}^l W_P$	Peierls energy
$W_i$	Elastic interaction energy between nucleus and dislocation
$W_i^{(1)}$	Misfit contribution to $W_i$
$W_i^{(2)}$	Modulus contribution to $W_i$
$W$	Ordering energy in Bragg-Williams theory
$\mathbf{x}_i$	Orthonormal co-ordinates
$\dot{x}$	Velocity
$x_0, x_\theta$	Mean distance between kinks in surface step
$x_A$	Atomic fraction of $A$
$x_B, x$	Atomic fraction of $B$

<i>Symbol</i>	<i>Meaning</i>
$x_a$	Fraction of active sites
$x^\alpha, x^\beta$	Compositions of phase boundaries
$x_\infty^\alpha$	Equilibrium $\alpha$ -phase composition for planar interface
$x_r^\alpha, x_r^\beta$	Equilibrium phase compositions for $\beta$ particles of radius $r$ in $\alpha$ matrix
$\Delta x_r^\alpha$	Change in solubility for $\beta$ particles of radius $r$
$x_{1c}, x_{2c}$	Compositions of coherent solubility gap
$x_0$	Mean composition of alloy
$x_w$	Fraction of "wrong" sites
$\mathbf{x}^{(0)}$	0-lattice vectors
$\mathbf{x}^{(02)}$	0-2 lattice vectors
$\mathbf{X}, X_{ij}$	Stress tensor
$X_{\text{crit}}$	Stress to cause slip in perfect crystal
$X_{ij}^I$	Stresses related to $e_{ij}^I$ by Hooke's Law
$X_{ij}^C$	Stress distribution in matrix from constrained transformation
$X_{ij}^T$	Stress distribution of transformed region in a constrained transformation
$X_p$	Peierls-Nabarro stress
$X$	Parameter in theory of interface controlled growth
$y_i$	Orthonormal co-ordinates
$y$	Normal separation of glide planes
$y^D$	Diffusion distance
$y$	Thickness of a plate
$y^{\alpha\beta}$	Thickness of lamellar aggregate of $\alpha+\beta$
$y_i$	Configurational co-ordinate
$y$	Unique semi-axis ellipsoid representing particles of varying shape
$Y$	Young's modulus
$Y'$	Elastic parameter
$Y$	Separation of dislocation lines in a boundary
$Y'$	Separation of dislocations in a boundary, projected normal to slip planes
$Y'$	Separation of disclination dipoles in boundary
$Y_{lm}(\theta, \phi)$	Spherical harmonic describing shape perturbation
$z_i$	Orthonormal co-ordinates
$z$	Co-ordination number
$z$	Number of interstitial sites per normal site
$z$	Mean number of neighbours in non-regular structure
$z_i$	Number of $i$ th neighbours
$z_{AB, i}$	Average number of $i$ th neighbours of opposite type
$z$	Position of interface
$z$	Exponent of $\Delta T^-$ in growth law
$\mathbf{Z}, Z$	Generalised force
$Z_n$	Quasi-steady distribution function giving number of embryos $Q_n$ in a metastable assembly
$Z_{n,t}$	Time dependent distribution function
$\alpha$	Constant in Morse potential

<i>Symbol</i>	<i>Meaning</i>
$\alpha$	Constant in interatomic potential
$\alpha$	Parameter in Peierls-Nabarro model
$\alpha_n$	Fourier coefficient in anisotropic elastic theory
$\alpha_+, \alpha_-, \alpha_0$	Probability of positive, negative or no kink at any site on a surface step
$\alpha^D$	Parameter used in the theory of nucleation on dislocations
$\bar{\alpha}$	Dimensionless supersaturation
$\bar{\alpha}_{\text{eff}}$	Dimensionless supersaturation; growth parameter
$\alpha_i, \alpha_j^*$	Growth parameters
$\alpha^m$	Growth parameter
$\alpha^B$	Parameter in theory of cellular growth
$\alpha^l$	Parameter in theory of cellular growth
$\alpha_i$	Orthogonal four-axis basis
$\alpha_i^*$	Orthogonal reciprocal basis
$\beta_n$	Fourier coefficient in anisotropic elastic theory
$\beta_c$	Critical wave number
$\beta$	Aspect ratio of ellipsoid of revolution
$\beta$	Parameter used in nearest neighbour model
$\beta^l, \beta'$	Parameters in theory of cellular growth
$\beta$	Wave vector of composition fluctuation
$\gamma_A, \gamma_B, \gamma^{\dagger}$	Activity coefficients
$\gamma, \gamma_1, \gamma_2$	Axial ratio
$\gamma$	Length of $\alpha_4$ axis
$\gamma$	Axial ratio of rectangular net
$\Gamma$	Coefficient of Gibbs-Thomson effect
$\Gamma_{\text{LP}}$	Lothe Pound correction factor
$\Gamma_Z$	Zeldovitch factor in nucleation rate
$\Gamma_{\text{Th}}$	Non-isothermal factor
$\Gamma$	Number of jumps per second made by a defect
$\Gamma_A$	Total number of jumps per second made by an $A$ atom
$\delta$	Parameter in average potential model
$\delta$	Thickness of a sheet or diameter of a wire
$\delta^B$	Thickness of a grain or interphase boundary
$\delta_{(r)}$	Amplitude of shape perturbation
$\dot{\delta}$	Growth rate of perturbation
$\delta_1$	Amplitude of shape perturbation of sphere
$\delta_{ij}$	Elements of substitution or identity matrix
$A$	Cubic dilatation
$\varepsilon$	Experimental activation energy per atom
$\varepsilon_0$	Difference in zero energy levels
$\varepsilon$	Activation energy for diffusion
$\varepsilon$	Fractional rate of change of lattice parameter with composition
$\varepsilon_i$	Principal strains of $e_{ij}^l$
$\varepsilon_{ijk}$	Alternating factor
$\eta$	Shape factor
$\eta^\beta, \eta^{\alpha\beta}, \eta^{\alpha S}$	Shape factors

<i>Symbol</i>	<i>Meaning</i>
$\eta_1$	Twinning direction
$\eta_2$	Conjugate twinning direction
$\theta$	Polar co-ordinate
$\theta$	Angle between surface step and close-packed direction
$\theta$	Parameter in average potential model
$\theta_i$	Dihedral angles between boundaries and/or surfaces
$\theta$	Misorientation of lattices across a grain boundary
$\theta$	Contact angle
$\theta_D$	Debye temperature
$\theta_+, \theta_-$	Angles of rotation
$\theta_j$	Angle between $i$ th and $i+j$ th atomic jumps
$X^\alpha$	Function of elastic properties
$\lambda$	Boltzmann variable in diffusion equation
$\lambda_{.A}$	Absolute activity
$\lambda$	Elastic constant of isotropic continuum
$\lambda, \lambda_n$	Coefficients in solution of diffusion equation
$\lambda_c$	Critical wave-length
$\lambda_i$	Eigenvalues of a matrix
$\lambda_i$	Principal shape deformations
$\mu$	Shear modulus of isotropic elastic continuum
$\nu_1$	Rate at which a site becomes a nucleus
$\nu$	Poisson's ratio
$\nu$	Characteristic vibration frequency
$\nu_i$	Normal atomic vibration frequencies
$\nu_i^0$	Normal frequencies for constrained assembly
$\nu_D$	Debye frequency
$\nu$	Normal to an invariant plane
$\xi$	Degree of advancement
$\xi$	Strain normal to martensite plate
$\xi$	Atomic position vector
$2E_1, 2E_2$	Binding energies (negatives of interatomic potentials) for nearest neighbour and second nearest neighbour atoms
$2E_{.AA}$	Binding energy of a nearest neighbour $A-A$ atom pair
$2E_{.AB}$	Binding energy of a nearest neighbour $A-B$ pair
$E$	Characteristic potential of nearest neighbour model
$E_0(r)$	Function in theory of conformal solutions
$E_{.AA}^*$	Characteristic energies in average potential model
$\pi_{.A12}$	Probability of an $A$ atom transferring from plane 1 to plane 2 in unit time
$\pi_{ \xi\eta}$	Probability of a vacancy moving from a plane of concentration $\xi$ to a plane of concentration $\eta$ in unit time
$\pi_{ \xi A}(\xi\eta)$	Probability of a vacancy on a plane of concentration $\xi$ exchanging places with an $A$ atom on a plane of concentration $\eta$ in unit time
$\varrho(r)$	Radial distribution function
$\varrho$	Parameter of average potential model
$\varrho$	Density

<i>Symbol</i>	<i>Meaning</i>
$\rho$	Mobile dislocation density
$\rho_1$	Density of dislocation loops
$\sigma(\mathbf{n})$	Surface free energy per unit area of surface with normal $\mathbf{n}$
$\sigma^{\alpha\alpha}$	Free energy per unit area of an $\alpha$ — $\alpha$ grain boundary
$\sigma^{sl}, \sigma^{\alpha\beta}$	Free energy per unit area of solid–liquid, alpha–beta interface
$\sigma_\infty$	Surface free energy per unit area of a planar interface
$\sigma_0$	Minimum value of $\sigma$ as an interface advances
$\sigma_{kl}$	Effective stress in pairwise model
$\sigma^j(x,y)$	Energy of generalised stacking fault
$\sigma^f$	Stacking fault (free) energy
$\sigma$	Short range order parameter
$\Sigma$	Reciprocal density of coincidence sites
$\tau_s$	Effective shear stress on slip plane
$\tau_0$	Parameter in empirical dislocation velocity equation
$\tau$	Time at which a region nucleates
$\tau$	Time to cross an activation barrier
$\tau''$	Mean time of stay at a dislocation
$\tau$	Relaxation time for establishment of a steady state nucleation rate
$\tau_0, \tau_1$	Coefficients in expansion of shear stress
$\varphi$	Angle giving deviation of tilt boundary from symmetrical position
$\varphi(z)$	Function giving periodic variation in energy with interface position
$\varphi_i(\alpha_i)$	Function used in theory of growth
$2\varphi$	Angle between $K_1$ and $K_2$ planes after twinning shear
$\varphi$	Angle between interface normal and rotation axis
$\varphi_D, \varphi'_D$	“Driving force” for diffusion
$\Phi$	Disregistry in Peierls-Nabarro theory
$\zeta$	Fractional amount of transformation; usually volume fraction transformed
$\zeta$	Width of dislocation
$\psi$	Angle between curve of constant composition and grain boundary
$\omega$	Strength of disclination dipoles
$\omega$	Wavenumber of shape perturbation
$\omega_c$	Critical wavenumber for stability
$\omega$	Rigid body rotation
$\Omega(N_A N_B N_{AB})$	Number of arrangements for given $N_{AB}$
$\Omega_a, \Omega_b$	Plane waves radiated from moving dislocation
$\Upsilon$	Velocity of an interface
$\Upsilon_c$	Velocity of planar interface under interface control
$\Upsilon_0$	Local velocity of interface

# *Author Index*

*Part I pp1–552*

*Part II pp553–1113*

- Aaronson, H. I. 449, 456, 479, 505–6, 518, 522, 527,  
714, 715, 717, 750 3, 755, 764, 793 5, 1089, 1092,  
1096, 1101
- Abdou, S. 772, 793
- Abe, T. 776, 777, 793
- Abell, J. S. 891, 892, 956
- Abraham, F. F. 441, 478
- Acuña, R. 735, 793
- Agarwal, S. 734, 793
- Agyeman, K. 1080, 1087
- Ahearn, P. J. 691, 698
- Ahlers, M. 1028, 1059
- Aizu, K. 1105, 1112
- Alder, B. J. 162, 167
- Alexander, K. B. 772, 793
- Allen, N. P. 912, 956
- Allen, S. 828 9, 831
- Allen, S. M. 211 12, 214, 233
- Almin, K. E. 1102, 1112
- Alstetter, C. 932, 957
- Altmann, S. L. 112, 167
- Altschuler, T. L. 891, 960
- Amelinckx, S. 251, 316, 322, 341, 376, 591, 619
- Anantharaman, T. R. 968, 990, 1077, 1087
- Anderson, J. C. 604, 620
- Anderson, W. A. 838 9, 857
- Andrade, E. N. 892, 956
- Androussi, Y. 901, 960
- Anthony, J. 382, 420
- Anthony, T. R. 413, 419
- Antonopoulos, J. G. 948, 956
- Aoki, M. 2, 22
- Ardell, A. J. 785 7, 793
- Arunchalaman, V. S. 886, 956
- Asaro, R. J. 267 8, 323
- Ashby, M. F. 354, 376, 922, 956
- Asimov, R. 786, 793
- Atkinson, C. 493–4, 506, 527, 756–8, 793
- Atwater, H. 655, 698
- Aubauer, H. F. 786, 793
- Aust, K. T. 350, 376, 847–9, 856–8
- Averbach, B. L. 225, 229, 233, 891, 960
- Avrami 838
- Avrami, M. 18, 20 1, 532, 535, 539, 541, 544, 546,  
550 2
- Ayers, J. D. 504, 527
- Babuska, I. 277, 322
- Backofen, W. A. 900, 960
- Bacon, D. J. 261, 267 9, 271, 323 4, 924, 949, 956,  
959, 1058, 1060
- Bagley, K. A. 684, 699
- Bailey, J. E. 835, 839 45, 857
- Bain, E. C. 548, 552, 1015, 1059
- Bainbridge, D. W. 837, 857
- Bak, P. 619
- Bakker, H. 397, 419
- Ball, C. J. 316, 323
- Balluffi, R. W. 2, 22, 113, 135, 142 5, 167, 170, 308,  
325, 366, 374, 376 7, 403, 412, 418 21, 922, 956
- Bannerjee, S. 1077, 1087
- Barbieri, A. 659, 698
- Barcello, G. 1107, 1112
- Bardeen, J. 250, 252, 323, 375, 391, 395, 397, 402 3,  
405, 419
- Bardsley, W. 676, 698
- Barford, J. 1096, 1101
- Barnes, R. S. 305, 326, 403, 419 20, 895, 957
- Barnett, D. M. 261, 267 8, 323, 476 8
- Barrett, C. S. 107, 110, 167, 890, 908, 944, 951, 959,  
983, 990
- Barry, D. E. 890, 958
- Bartlett, A. F. 891, 960
- Bartlett, J. H. 138, 167
- Basinski, A. S. 1045 6, 1059

- Basinski, Z. S. 119-20, 126, 167, 266, 278, 280, 290,  
302, 311, 313, 323, 325, 366, 376, 984, 988-90,  
1044, 1046, 1052, 1112
- Baskes, M. I. 2, 22
- Bass, J. 143, 145, 168
- Bassett, G. A. 591, 598, 619
- Bauerle, J. E. 141 2, 144, 167, 169
- Bauser, E. 592, 619
- Beck, L. H. 227, 233
- Beck, P. A. 841, 846, 847, 850-2, 857, 858
- Becker, J. H. 486, 527, 704, 706 7, 716
- Becker, R. 398, 420, 422, 432, 436 9, 441 2, 445, 478,  
548, 637, 698, 721, 739, 793
- Beeler, J. R. 125, 131 2, 138, 167 8, 353 4, 377
- Begley, R. T. 891, 960
- Bell, R. L. 913, 956
- Bell, T. 751, 794
- Bellemans, A. 199, 234
- Bendersky, L. 617, 619
- Benedicks, C. 1102, 1112
- Benet, H. P. 695, 700
- Bennema, P. 579, 592, 619, 621
- Bennett, M. R. 616, 619
- Bensimon, D. 662, 698
- Benson, G. C. 423, 478
- Berghezan, A. 845, 858
- Bernal, J. D. 162 3, 166 7, 354
- Berner, R. 299, 325
- Bernstein, H. 178, 233
- Bernstein, I. M. 891, 960
- Bethe, H. 153 5, 167, 186, 193, 218, 220 4, 233
- Bethe, H. 591, 608, 619 20
- Bever, M. B. 110, 169, 832, 857 8, 1102, 1113
- Bevis, M. 292, 325, 867, 881, 886, 888, 891, 896, 899,  
903, 955 7, 959
- Bhadeshia, H. K. D. H. 1089 90, 1092 3, 1097,  
1099 101
- Bhattacharyya, S. K. 703, 713, 716
- Bianchi, G. 135, 167
- Bibby, M. J. 702, 716
- Bibring, H. 965, 990
- Bienenstock, A. 426, 431, 442, 479
- Biggs, W. D. 912, 956
- Bilby, B. A. 15, 21, 32, 78, 168, 245, 250, 266, 270,  
289, 291, 293 4, 300, 316 17, 319 20, 322 3, 343,  
365 6, 370, 373, 376, 506 8, 527, 544 5, 860, 867,  
874, 881, 908, 932 3, 935 8, 940, 956, 957, 961,  
963 4, 990, 993, 995, 1003, 1025, 1035, 1048,  
1059 60
- Biloni, H. 675 6, 698
- Binder, K. 558, 619
- Birnbaum, H. 931, 956
- Biscondi, M. 353 5, 377
- Bishop, G. H. 351, 376
- Blackburn, L. D. 110, 167
- Blech, I. 165, 170
- Blewitt, T. H. 890, 908, 956
- Bloem, J. 596-7, 619
- Boas, G. 803, 817, 971, 973
- Boazs, W. 971 3, 991
- Boivin, P. 901, 960
- Bolling, G. F. 502 3, 527, 631, 658, 698, 810, 817,  
852, 855, 857, 953, 956
- Bollman, W. 841, 857, 1059
- Bollmann, W. 351, 366, 368, 370 7, 923, 956, 958,  
1048
- Bonfiglioli, A. 735, 793
- Bonnet, R. 1049, 1059
- Bonnissent, A. 558, 619
- Bontinck, W. 251, 322
- Boos, J. Y. 353 5, 377
- Borelius, G. 721, 767, 769, 780, 793
- Bormann, R. 1081, 1087
- Born, M. 115 16, 131, 161, 167, 277, 327, 376
- Boswell, P. G. 634, 698
- Boucher, N. A. 891, 912, 953, 956
- Bourquin, R. D. 353 4, 377
- Bowen, D. K. 120, 170, 250, 278, 311, 313, 324 5,  
973, 990
- Bower, T. F. 679 80, 691 2, 698
- Bower, T. P. 691, 698
- Bowles, J. S. 25, 78, 965, 990, 993, 1018 24, 1028 9,  
1031, 1034, 1036 8, 1043, 1045 7, 1059 60
- Bradley, J. R. 753, 755, 793
- Bradshaw, F. J. 141, 167
- Bragg, W. L. 155, 167, 216 27, 233, 336, 376
- Brailsford, A. D. 268, 323, 786, 793
- Braisaz, T. 948, 957
- Brakss, N. 684, 701
- Bramfitt, B. L. 810, 817
- Brandon, D. G. 351, 376
- Brandreth, M. 763, 793
- Brandt, W. H. 514 16, 527
- Braumann, F. 773, 781, 794
- Breedis, J. F. 1019, 1022, 1059
- Bridgeman, P. W. 650, 698
- Bristowe, P. D. 356, 376, 919, 921, 932, 957
- Brody, H. D. 692, 698
- Brook, R. 1073 4
- Brooks, H. 329, 337, 363, 376
- Bross, H. 130 1, 170
- Brown, D. 810 1, 817
- Brown, L. M. 267 8, 323
- Brown, T. E. 652, 701
- Brown, W. B. 199, 233
- Buckel, W. 1078, 1087
- Buchler, W. J. 1102, 1112
- Buenger, M. J. 6, 21, 653, 673, 698
- Bufalini, P. 347, 377
- Buff, F. P. 423, 441, 478 9

- Buhrman, R. A. 620  
 Bullough, R. 32, 78, 245, 262, 266 7, 278, 317, 323, 507 8, 527, 544 6, 552, 901, 905, 957, 993, 1059  
 Bunshah, R. F. 977, 990  
 Burden, M. H. 679 80, 698  
 Burger, G. 147, 170  
 Burgers, J. M. 239, 258, 266, 323, 336, 376  
 Burgers, W. G. 704 5, 716 7, 768, 795, 840, 850 1, 857, 1059  
 Burkart, M. W. 982, 984, 990, 1102, 1112  
 Burke, J. E. 850, 852, 855, 857  
 Burns, F. P. 826, 831  
 Burton, J. A. 669, 698  
 Burton, J. J. 143 5, 167, 478  
 Burton, W. K. 148, 150 1, 153 4, 156, 159 60, 167, 562 3, 576 7, 579, 581, 583, 586, 590, 619  
 Butcher, B. R. 1047, 1059  
 Butler, E. P. 767 8, 772, 793, 796  
 Bystrom, A. 1102, 1112
- Cabrera, N. 148, 150 1, 153 4, 156, 159 60, 167, 562, 568, 577, 579, 584, 588, 619  
 Cagle, F. W. 706, 716  
 Cahn, J. W. 8, 22, 165 70, 184 5, 211 14, 233 4, 294, 323, 391, 398, 400, 418, 420, 458 60, 472 8, 481 2, 488, 501 3, 505, 510, 515 16, 518 23, 525 7, 530, 532 6, 550 2, 557, 619, 657, 699, 722, 724, 727 8, 742, 770, 787, 793, 796, 802, 806, 808 9, 811 2, 815, 817, 828 9, 831, 835 6, 850, 857, 885 6, 894 5, 957, 959  
 Cahn, R. W. 55, 78, 834, 840, 847, 857 8, 901, 906, 909, 913, 956 7, 1077, 1079 80, 1084, 1087  
 Calka, A. 1080, 1087  
 Callaway, J. 112, 168  
 Calnan, E. A. 750, 793  
 Calvert, J. 774, 793  
 Campbell, D. R. 414, 420  
 Cantor, B. 634, 697, 700 1, 716 7, 756, 794, 1076, 1078, 1087  
 Carrington, W. 316, 323, 341, 376  
 Carslaw, H. S. 388, 420, 450, 478  
 Castaing, R. 776, 793  
 Cech, A. E. 977 9, 990, 1071  
 Cech, R. E. 627, 630, 638, 698, 701, 1074  
 Celli, V. 277, 323  
 Chadwick, G. A. 631, 633 4, 642 3, 651, 655, 687, 698 701  
 Chakraborty, S. B. 893 4, 957  
 Chalmers, B. 328, 351, 376, 567, 619, 647, 651 3, 655, 663, 672, 674, 690, 693, 695, 698 701  
 Champier, G. 135, 167  
 Chang, L. C. 984, 1102, 1112  
 Chang, R. 270, 278, 323  
 Chapon, C. 608, 619
- Charbnau, H. P. 135, 168, 356, 376  
 Chattopadhyay, K. 617, 619, 1085, 1088  
 Chaudhari, P. 356, 376  
 Chaudron, G. 695, 699  
 Cheetham, D. 810 1, 817  
 Chen, H. S. 1080, 1088  
 Chik, K. P. 142 3, 145, 168  
 Childs, V. A. 633, 699  
 Chin, G. Y. 890 1, 912, 944 5, 956, 958, 960  
 Chipman, W. B. 227, 233  
 Chiswick, H. H. 901, 906, 958  
 Chopra, K. L. 110 11, 168  
 Chopra, M. 683, 699  
 Chou, Y. T. 270, 323  
 Christian, J. W. 6, 8, 15, 21 2, 124 5, 168, 256, 277, 279 80, 289 90, 293, 311, 322 3, 325, 354, 364, 366, 370, 376, 470, 478, 533, 552, 702, 716, 860, 877, 885 6, 890 1, 893, 895, 901, 909, 912, 919, 922, 924, 934, 936 7, 947, 951, 953, 956 7, 959 60, 961, 963 4, 965 8, 973, 984, 988 9, 994 5, 998, 1033, 1040, 1042 3, 1044 6, 1048, 1052, 1055, 1058 60, 1089, 1100 2, 1112  
 Churchman, A. T. 905, 957  
 Clapp, C. 828, 831  
 Clapp, P. C. 15, 21 2, 211, 214, 233  
 Clareborough, L. M. 305, 307 8, 323, 325, 832, 857  
 Clark, H. M. 6, 22  
 Clark, J. B. 768, 794  
 Classen, W. A. P. 596 7, 619  
 Clemm, P. J. 456 8, 478  
 Clews, C. J. B. 793, 750  
 Cline, H. E. 643, 700  
 Clougherty, E. V. 109, 169  
 Coates, D. E. 755, 793  
 Cochardt, A. W. 271 2, 323  
 Cockayne, D. J. H. 299, 323  
 Cohen, E. R. 430 1, 479  
 Cohen, J. B. 893 4, 958  
 Cohen, M. 8, 15, 21 2, 110, 167, 229, 233, 271, 324, 889, 891, 894, 919, 957 8, 980 3, 985 7, 990 1, 1020, 1023, 1025, 1049, 1060, 1062, 1064 8, 1073, 1075  
 Cohen, M. L. 113, 168  
 Cole, G. S. 675, 699  
 Coleman, B. D. 102 3, 105, 420  
 Coleman, R. V. 588, 612, 619, 621  
 Coll, J. A. 894, 957  
 Collins, R. 163, 168  
 Compaan, K. 397, 420  
 Conrad, G. P. 899, 891, 959 60  
 Conrad, H. 280, 323  
 Cook, H. E. 400, 420, 725, 729, 734 6, 793, 1068, 1075  
 Cook, J. M. 1110, 1112  
 Corbett, J. W. 147, 168

- Cormia, R. L. 162, 164, 170, 627, 701  
 Cotterill, R. M. J. 131–2, 165, 168–9, 278, 323 4  
 Cottrell, A. H. 128, 133, 168–9, 178, 233, 250, 263,  
 269–72, 281, 300–2, 305, 311, 323, 472–3, 506 8,  
 527, 544 5, 840, 857, 932, 935, 938 40, 957  
 Cottrell, S. A. 962, 991, 1093, 1101  
 Couchman, P. R. 329, 376  
 Couling, S. L. 881, 898, 900, 957  
 Courant, R. 450, 478  
 Cowley, J. M. 225 6, 233, 821, 831  
 Cox, J. J. 912, 957  
 Cracknell, A. P. 112, 168  
 Crank, J. 388, 420  
 Crocker, A. G. 42, 51, 78, 354, 356, 370, 376 7, 860,  
 865, 867, 874, 880, 881, 886, 888, 891, 892, 896,  
 898, 899, 903, 906, 907, 918, 919, 921, 922, 932,  
 956 7, 1002, 1003, 1025, 1035, 1039, 1041, 1047,  
 1060  
 Crussard, C. 80, 94, 271, 323, 834, 857, 990, 991  
 Cwilong, B. M. 628, 699  
 Czocharalski, T. 650, 699
- Dahl, R. E. 353 4, 377  
 Dahlberg, E. P. 900, 960  
 Dahmen (1990) 901  
 Damask, A. C. 131, 168  
 Daniel, V. 721, 793  
 Darken, L. S. 391, 394 5, 402 3, 406 7, 420, 517,  
 527, 800, 817  
 Darwin, C. G. 652, 699  
 Dash, W. C. 250, 323, 655, 699, 761, 793  
 Davenport, E. S. 548, 552  
 Davies, A. H. 1087  
 Davies, C. K. L. 786, 793  
 Davies, H. A. 1080, 1088  
 Davies, P. W. 843, 846, 857  
 Davies, V. De L. 687, 699  
 Davis, R. S. 656, 699  
 Daw, M. S. 2, 22  
 Dawson, D. B. 441, 478  
 Day, M. G. 641, 686 7, 699  
 De Fontaine, D. 214, 233, 725, 735 6, 793, 827, 828,  
 831  
 De Groot, S. R. 102, 104 5  
 De Haan, S. W. H., 579, 619  
 De Hossen, J. Th. M. 1082, 1087  
 De Jong, M. 131, 168, 305, 307, 323, 325  
 De Sorbo, W. 770, 782 3, 794  
 De Wit, R. 266 8, 295, 323  
 Dean, F. H. 670, 699  
 Debye, P. 80, 92 4, 107 10, 223, 411, 426  
 Decker, B. F. 839, 857  
 Dehlinger, U. 721, 793, 1065 6, 1075  
 Dehoff, R. T. 755, 794
- Dekeyser, W. 251, 322  
 Delaey, L. 755, 795, 1061, 1102, 1112  
 Delamotte, E. 932, 957  
 Delves, R. T. 678, 699  
 Detert, K. 849, 858  
 Devonshire, A. F. 164, 169  
 Dickey, J. E. 277, 324  
 Dienes, G. J. 129, 131, 138, 167 8, 170, 822, 824–5,  
 831  
 Dietze, H. D. 279, 298, 323  
 Dimitrov, O. 847, 857  
 Dingley, D. J. 374, 377  
 Dippenaar, R. J. 805, 817  
 Dobbins, R. A. 559, 621  
 Dobson, P. S. 308, 310, 324  
 Doherty, R. D. 756, 794, 850, 857  
 Doherty, R. 656, 680, 699  
 Doi, M. 785, 786, 788 91, 792, 794  
 Dollins, C. C. 476 8  
 Domian, K. A. 755, 793  
 Donald, I. W. 1080, 1087  
 Donnadieu, P. 618, 619  
 Donth, H. 280, 325  
 Doremus, R. H. 490, 527, 542 4, 552, 744, 745,  
 746 7, 794  
 Döring, W. 422, 432, 436 9, 441 2, 445, 478  
 Dörner, W. 1080, 1087  
 Double, D. 687, 688, 699  
 Dove, D. B. 860, 863, 864, 866, 958  
 Doyama, M. 131 2, 143, 168, 278, 323 4  
 Dubois, J. M. 1080, 1082, 1087  
 Duesbery, M. S. 119 20, 126, 167 8, 278, 311, 313,  
 323 4  
 Duff, R. H. 110 11, 168  
 Duflos, B. 1076, 1087  
 Dunn, C. G. 854, 858  
 Dunne, C. G. 837, 857  
 Dunne, D. F. 1021, 1023, 1060  
 Dunne, D. P. 1028, 1059  
 Dunning, W. J. 426, 478, 708, 716  
 Duwez, P. 631, 699, 702, 703, 714, 716, 1076, 1078,  
 1080, 1087  
 Dyson, B. F. 382, 420
- Ebisuzaki, Y. 409, 420  
 Edington, J. W. 300, 305, 324  
 Edmonds, D. V. 1089, 1090, 1092, 1101  
 Edmondson, B. 891, 960, 989, 990  
 Edmunds, G. H. 908, 958  
 Edmunds, I. G. 821, 831  
 Edwards, E. H. 376 7  
 Edwards, O. S. 125, 168  
 Efsic, E. J. 1021, 1023, 1060  
 Egami, T. 1083, 1087

- Eggington, A., 558, 619  
 Einstein, A. 80, 92, 385, 395, 411, 420, 506  
 Elbaum, C. 567, 619, 651, 699  
 Ellwood, E. C. 684, 699  
 Enomoto, M. 753, 756, 786, 7, 794  
 Entwisle, A. R. 1014, 1025, 1060, 1071, 1075  
 Esaka, H. 683, 699  
 Escaig, B. 305, 324  
 Eshelby, J. D. 135, 168, 201, 203–5, 233, 261, 267,  
 270, 274, 276, 283–5, 324, 466, 468, 476–8  
 Evans, A. G. 280, 324  
 Evans, H. 820, 831  
 Evans, J. L. 146, 168  
 Ewald, A. W. 704, 705, 716  
 Ewing, H. H. 627, 699  
 Eyre, B. L. 146, 168  
 Eyring, H. 79–80, 85, 91, 94, 483, 706, 716  
 Faridi, B. A. 354, 376  
 Farkas, L. 436, 439, 479  
 Fearon, E. O. 891, 957  
 Feder, J. 430, 440, 451, 479  
 Feder, R. 135, 168, 9, 426, 479, 826, 827, 831  
 Fehling, J. 629, 699  
 Feltham, P. 852, 857  
 Finney, J. L. 163, 4, 168  
 Finnis, M. W. 2, 22, 113, 168  
 Fischer, M. 1080, 1087  
 Fisher, J. C. 414, 15, 418, 420, 434, 442, 5, 456, 8,  
 478, 9, 513, 517, 18, 527, 782, 794, 807, 8, 817,  
 855, 7, 1064, 5, 1067, 1070, 2, 1074, 5  
 Fisher, R. M. 517, 527, 800, 817, 821, 831  
 Fisher, X. 558, 619  
 Flemings, M. C. 679, 688, 691, 2, 693, 694, 698, 700  
 Fletcher, N. H. 356, 377  
 Flinn, P. A. 8, 22, 229, 233  
 Flood, H. 441, 479, 559, 622  
 Flynn, C. P. 136, 143, 168  
 Foreman, A. J. E. 262, 268, 276, 7, 323, 4  
 Forewood, C. T. 307, 324  
 Fortes, M. A. 349, 377  
 Forty, A. J. 591, 619  
 Fournelle, R. A. 768, 772, 794  
 Fowler, C. M. 891, 960  
 Fowler, R. H. 221, 233  
 France, L. K. 313, 324  
 Frank, F. C. 6, 22, 40, 46, 78, 106, 120, 1, 123, 126,  
 150, 152, 4, 156, 158, 9, 161, 2, 165, 8, 244, 248, 52,  
 257, 268, 284, 287, 294, 7, 300–1, 304, 7, 310,  
 314, 16, 324, 341, 3, 350, 377, 382, 413, 420, 463,  
 472, 488, 527, 567, 568, 570–1, 573, 578, 580,  
 585, 589, 591, 592, 619–20, 651, 654, 663, 671,  
 699, 761, 794, 906, 911, 927, 957, 1023, 1048,  
 1049, 50, 1060  
 Frankl, D. 599, 605, 620  
 Freedman, J. F. 181, 189, 198, 206, 233  
 Freise, E. J. 775, 794, 900, 902, 957  
 Frenkel, J. H. 146, 149, 151, 168, 424, 426, 429–30,  
 436, 449, 479  
 Freundlich, H. 184, 233, 427  
 Fridberg, J. 806, 817  
 Friedel, J. 205–6, 233, 257, 268, 295, 324  
 Frois, C. 847, 857  
 Frye, J. H. 811, 817  
 Fuchs, K. 116–17, 168  
 Fujita, H. 890, 913, 944–5, 957–8  
 Fullman, R. L. 855–7  
 Fumi, F. G. 130, 168  
 Fuoss, R. M. 392, 420  
 Furth, R. 118, 168  
 Furukawa, K. 143, 161, 168, 170  
 Gallagher, P. C. J. 304, 324, 945, 957  
 Galligan, J. M. 146, 168  
 Gallot, J. 127, 170  
 Garcia-Garcia, L. M. F. 890, 913, 930, 958  
 Gardner, W. E. 112, 169  
 Garwood, R. D. 715, 717  
 Gaskell, P. H. 1082, 1087  
 Gaunt, P. 702, 716, 862, 900, 959, 965, 967, 990,  
 1040, 1042, 3, 1055, 1060  
 Gavazza, S. D. 267, 8, 323  
 Gay, P. 834, 857  
 Gehlen, P. C. 125, 131, 168, 278, 9, 324  
 Geisler, A. H. 774, 794  
 Gensamer, M. 896, 959  
 George, B. 135, 169  
 Gerold, V. 776, 794, 953, 957  
 Gibbs, J. W. 1, 4, 10, 101, 107, 182, 4, 200, 233,  
 393, 4, 398, 427, 487, 490, 495, 9, 501, 4, 512, 516,  
 523, 526, 557, 620  
 Gibson, J. B. 131, 133, 168  
 Giessen, B. C. 110, 169, 1080, 1087  
 Gilbert, A. 710, 711, 717, 1062, 1075  
 Gilman, J. J. 285, 6, 324  
 Gilmer, G. H. 562, 3, 578, 9, 586, 588, 620, 622  
 Girifalco, L. A. 118, 131, 168  
 Glasstone, S. 79, 94  
 Gleiter, H. 328, 354, 376, 7, 611, 614, 616, 620  
 Glicksman, M. 633, 699  
 Glicksman, M. E. 504, 527, 631, 656, 60, 663,  
 699, 700, 786, 7, 795  
 Glocker, R. G. 781, 794  
 Glossop, A. B. 228, 233  
 Glover, S. G. 988, 989, 990  
 Goland, A. N. 131, 133, 168  
 Gomer, R. J. 612, 620  
 Gomez-Ramirez, R. 477, 479

- Gomsall, S. Y. 629, 699  
 Goo, E. 893, 894, 957  
 Goodenow, R. H. 1096, 1101  
 Goodrew, P. J. 308, 324  
 Gordon, P. 838, 858  
 Goswami, R. 634, 699  
 Gough, H. J. 901, 960  
 Goux, C. 353 5, 377  
 Graf, L. 663, 699  
 Graham, C. D. 847, 857  
 Graham, L. J. 278, 323  
 Grandqvist, C. G. 615, 620  
 Grant, N. J. 110, 169  
 Gratias, D. 165, 170, 375, 377, 918, 957  
 Gray III, G. T. 890, 960  
 Green, H. S. 161, 167 8  
 Green, M. L. 889, 894, 957  
 Green, R. E. 847, 857  
 Greenwood, G. W. 785, 794  
 Greer, A. L. 1077, 1079, 1080, 1084, 1087  
 Greninger, A. B. 702, 709, 717, 961, 979, 990 1, 992, 1018 20, 1021, 1023, 1025, 1060, 1098, 1101  
 Grimmer, H. 374, 377  
 Gripshover, R. J. 145, 168  
 Groen, L. J. 704, 705, 716 7  
 Groves, P. P. 269, 324  
 Grunbaum, B. 618, 620  
 Guedo, J. Y. 893 4, 957  
 Guggenheim, E. A. 10, 22, 179, 187, 193, 195 7, 216, 221, 225, 227, 233  
 Guimier, A. 893 4, 957  
 Guimier, A. 233, 776, 778 80, 794, 796  
 Gulliver, C. H. 666, 699  
 Gupta, D. 414, 420  
 Gupta, S. P. 816, 817  
 Gurney, R. W. 129, 169  
 Gust, W. 771, 772, 794  
 Gutmanas, E. Y. 611, 620  
 Guttman, L. 226, 233  
 Guy, A. G. 8, 22  
 Guyoncourt, D. M. M. 891, 892, 918, 957  
 Guyot, P. 356, 377
- Haasen, P. 324, 890, 908, 944, 957, 1082, 1085, 1088  
 Hafner, J. 113, 168  
 Hagel, W. C. 515, 518, 527, 802, 806, 811, 816, 817  
 Hale, K. F. 316, 323, 341, 376  
 Hales, R. 310, 324  
 Hall, E. O. 705, 717, 860, 898, 903, 957  
 Ham, F. S. 488, 490, 494 5, 498, 501, 503, 527, 542 5, 552, 604, 620, 747, 794
- Handler, G. S. 128, 169, 380, 420  
 Handscomb, D. C. 46, 78  
 Hansen, M. 804, 817  
 Hansson, B. 895, 957  
 Hardy, A. H. 207-8, 233, 611, 620  
 Hardy, H. K. 741, 794  
 Harker, D. 839, 857 8  
 Harnecker, K. 908, 957  
 Harper, S. 544 5, 552  
 Harris, J. E. 310, 324  
 Harris, S. 566, 620  
 Harris, W. F. 923, 958  
 Harris, W. T. 1073, 1075  
 Harrison, L. G. 414, 418, 420  
 Harrison, W. A. 112 13, 168  
 Harsdorff, M. 608, 620  
 Hart, E. W. 417 18, 420, 783, 794  
 Hartley, C. S. 903, 959  
 Hasegawa, D. R. 1080, 1087  
 Hasegawa, S. 1080, 1087  
 Hasson, G. 353 5, 377  
 Hauffler, R. 228, 234  
 Haven, Y. 397, 420  
 Hawbolt, E. B. 703, 717  
 Hawkins, M. J. 1096, 1101  
 Haworth, W. C. 714, 717  
 Haynes, R. 802, 803, 817  
 Hayzelden, C. 716, 717, 965, 990, 1076, 1077, 1078, 1087  
 Hazzledine, P. M. 942, 950 1, 958 9  
 Head, A. K. 268 9, 295, 308, 313, 323 4  
 Heal, T. J. 741, 778, 794 5  
 Hedges, J. M. 316, 324  
 Hehemann, R. F. 1089, 1092, 1096, 1101  
 Heidenreich, R. D. 296, 298, 324  
 Heine, V. 113, 119, 127, 168  
 Hellawell, A. 641, 670, 673, 679, 686 7, 688, 690, 691, 699 700, 763, 793, 796  
 Henley, C. L. 1085, 1087  
 Henry, C. 608, 620  
 Henry, C. R. 608, 619  
 Hensel, F. 689, 699  
 Herbert, P. M. 690, 691, 699  
 Herbeuval, I. 353 5, 377  
 Herman, H. 734, 793  
 Herring, C. 156 8, 168, 250, 252, 323, 328 9, 331 2, 375, 377, 391, 395, 397, 403, 405, 419  
 Hess, J. B. 983, 990  
 Heumann, T. 403, 420  
 Hibbard, W. A. 837, 857  
 Hilbert, D. 450, 478  
 Hildebrand, J. H. 166, 168, 187  
 Hill, J. K. 774, 794  
 Hill, P. G. 441, 478 9

- Hillert, M. 8, 22, 178, 233, 400, 420, 498, 501, 505, 509, 516, 523, 526, 7, 722, 794, 803, 806, 817, 1062, 1063, 1075, 1089, 1101
- Hilliard, J. E. 166, 9, 184, 5, 233, 400, 420, 557, 619, 722, 723, 727, 729, 732, 4, 742, 793, 5
- Hillig, B. 578, 9, 620
- Hillig, W. B. 451, 479, 484, 505, 527
- Hillig, W. 713, 717
- Hilsch, R. 1078, 1087
- Hinde, R. M. 821, 831
- Hiraga, K. 1086, 1087
- Hirotsu, Y. 750, 794
- Hirsch, P. B. 142, 169-70, 254, 301, 304, 5, 307, 311, 316, 323-5, 834, 835, 840, 841, 842-5, 857
- Hirth, J. P. 185, 234, 268, 271, 280, 285, 324, 366, 374, 377, 440, 479, 569, 572, 577, 620, 1, 938, 939, 940, 942, 958
- Hitzenberger, C. 967, 1055, 1060, 1067, 1075
- Ho, P. S. 414, 420
- Hobstetter, J. N. 739, 794
- Hockstra, S. 1099, 1101
- Hoffman, R. E. 415, 16, 421
- Hohenberg, P. 2, 22
- Holden, J. 910, 958
- Hollomon, J. H. 434, 441, 479, 631, 641, 699, 782, 794, 990, 977, 978, 1065, 1071, 1074, 5
- Holmes, E. L. 849, 852, 853, 857
- Holt, D. B. 266, 325
- Honda, R. 891, 960
- Honeycombe, R. W. K. 805, 817
- Hong, S. I. 953, 958
- Honma, T. 965, 991
- Hono, K. 1085, 1087
- Hornbogen, E. 771, 795, 816, 817, 891, 1102, 1112
- Horvath, J. 1080, 1087
- Horvay, G. 488, 501, 3, 527
- Horvay, G. A. 657, 699
- Houska, C. R. 968, 990
- Howard, R. E. 103, 105, 382, 397, 406, 420
- Howe, J. M. 765, 6, 794, 1067, 1075
- Hu, H. 839, 840, 847, 856, 857, 8
- Huang, K. 115, 167
- Huang, S. C. 656, 663, 699
- Hubbard, J. N. 708, 717
- Huber, R. W. 21, 2
- Hull, D. 15, 22, 305, 311, 324, 326, 715, 717, 891, 911, 958, 960, 1, 990
- Hull, F. C. 801, 806, 808, 812, 817
- Hullett, L. D. 573, 620
- Hultgren, A. 510, 513, 527
- Humble, P. 307, 324
- Hume-Rothery, W. 107, 169, 172, 207, 227, 233
- Hunt, D. 684, 686, 699
- Hunt, J. D. 525, 527, 646, 656, 662, 676, 679, 80, 682, 3, 685, 698, 700
- Hunter, G. A. 687, 699
- Huntington, H. B. 76, 78, 129-30, 132, 3, 138-9, 169, 206, 277, 324, 381, 2, 420, 1
- Husband, J. N. 533, 552
- Huston, E. L. 726, 730, 731, 2, 794
- Hutchings, P. J. 892, 956
- Hutchinson, T. E. 606, 621
- Hyman, E. D. 749, 794
- Indenbom, V. L. 267, 324
- Ingersoll, L. R. 624, 700
- Ingle, K. W. 356, 377
- Ino, S. 560, 614, 620
- Inokuti, Y. 1077, 1087
- Inoue, A. 1080, 1087
- Irvine, K. J. 1089, 1092, 1101
- Isaitshev, I. 1035, 1060
- Isebeck, K. 147, 170
- Isenberg, I. 116, 169
- Ishimasa, T., 618, 620
- Israel, A., 593, 620
- Ivantsov, G. P. 501, 3, 527, 657, 699
- Ives, M. B. 573, 620
- Jack, K. H. 742, 794
- Jackson, J. J. 143, 169
- Jackson, K. A. 155, 167, 169, 525, 527, 562, 564, 620, 650, 679, 684, 686, 699, 700
- Jackson, P. J. 564, 611, 621
- Jacquet, P. A. 835, 857
- Jacucci, G., 558, 620
- Jaeger, H. L. 441, 479
- Jaeger, J. C. 388, 420, 450, 478
- Jaffee, R. I. 125, 131, 168
- Jaffrey, H. 655, 700
- Jagodzinski, H. 782, 794
- Janot, C. 135, 167, 169
- Jaswon, M. A. 61, 78, 276, 7, 324, 860, 863, 4, 866, 958
- Jena, A. K. 110, 169
- Jenkins, M. L. 299, 323
- Jensen, E. J. 165, 169
- Jesser, W. A. 329, 376
- Johannson, C. H. 228, 233
- Johnson, R. A. 119, 20, 126, 7, 131, 4, 138, 9, 144, 167, 169, 278, 312, 353
- Johnson, W. A. 18, 22, 393, 404, 406, 420, 530, 536, 7, 552
- Johnson, W. C. 456, 479
- Johnson, W. L. 1080, 1088
- Johnston, W. G. 285, 324

- Jones, D. R. H. 631, 700  
 Jones, F. W. 820, 831  
 Jones, G. J. 505, 527, 755–7, 794  
 Jones, H. 117, 169  
 Jøssang, T. 267–8, 324
- Kalos, M. H. 558, 619  
 Kantrowitz, A. 440, 450–1, 479  
 Karnthaler, H. P. 953, 957, 965–6, 967, 968, 991, 1055, 1060–1, 1067, 1075  
 Kaspar, J. S. 120–1, 161–2, 165, 168–9  
 Kass, W. J. 409, 420  
 Kattamis, T. Z. 691–2, 700  
 Katz, J. L. 430–1, 479, 559, 620  
 Kauffman, J. W. 142, 144, 169  
 Kaufman, L. 8, 22, 109–10, 167, 169–70, 178, 233, 980, 983, 990, 1062, 1066, 1068, 1075  
 Kawasaki, K. 786 7, 794–5  
 Kear, B. H. 893, 894, 958  
 Keating, D. T. 227, 233  
 Kellar, J. N. 834, 858  
 Keller, K. W. 591, 620  
 Kelly, A. 834, 857, 900, 902, 908, 957 8  
 Kelly, P. M. 1028 30, 1052 3, 1057, 1060  
 Kelton, K. F. 1083, 1085, 1088  
 Kelvin, Lord 333  
 Kessler, D. 659, 662, 700  
 Khachatryan, A. G. 214, 233, 792, 794, 827 8, 831  
 Khaikin, S. E. 695, 700  
 Khandros, L. C. 979, 991  
 Khoo, T. L. 136, 169  
 Khoshnevisan, M. 145, 168  
 Kiefer, B. 228, 234  
 Kiho, H. 866, 903, 958  
 Kikuchi, R. 196, 233  
 Kilgour, D. M. 163, 170  
 Kim, W. T. 634, 700  
 Kim, Y. H. 1080, 1088  
 Kimura, H. 141 2, 169  
 King, A. D. 751, 794  
 King, A. H. 890, 944, 957  
 Kino, T. 143–5, 169  
 Kinsman, K. R. 505 6, 527, 752, 794, 1089, 1101  
 Kirkaldy, J. S. 103, 105, 392, 420, 516, 523, 527 8, 752, 795, 811, 817  
 Kirkendall, E. O. 250, 382, 390 1, 394, 403, 407, 421  
 Kirkwood, J. G. 161, 169, 197, 233, 423, 479  
 Klement, M. 1076, 1088  
 Kléman, M. 257, 324  
 Knabbe, E. A. 608, 620  
 Knapp, H. 1065 6, 1075  
 Knodler, A. 780, 794  
 Knowles, K. M. 919, 957  
 Ko, T. 962, 989, 991, 1093, 1101
- Koch, E. F. 746–7, 794  
 Koehler, J. S. 131, 136, 141–5, 147, 167–70, 254–5, 258, 260, 267–8, 295, 307, 323–5  
 Kohn, W. 2, 22  
 Kohnke, E. E. 704, 716  
 Koistinen, D. P. 1072, 1075  
 Kolmogorov, A. N. 18, 22  
 Komen, Y. 374, 376  
 Kondo, K. 316  
 Konobeevsky, S. T. 420  
 Koptsik, V. A. 376–7  
 Koskenmaki, D. C. 1085, 1088  
 Koss, R. C. 145, 169  
 Kossel, W. 561, 562, 620  
 Köster, W. 773, 780, 781, 794  
 Kotler, G. R. 658, 700  
 Kramer, T. J. 680, 700  
 Krauklis, L. P. 1024, 1060  
 Krauss, G. 750, 795  
 Krisement, O. 1063, 1075  
 Kristensen, W. D. 165, 169  
 Krohn, M. 608, 620, 622  
 Kronberg, M. L. 289, 324, 345, 377, 846 7, 858, 916, 958  
 Kroupa, F. 266 7, 277, 322, 324  
 Kröner, E. 261, 266, 316, 324, 466, 479  
 Kuhlmann-Wilsdorf, D. 141 2, 169, 252, 277, 279, 300, 306 7, 324, 329, 376  
 Kuhrt, F. 424, 426, 479  
 Kulin, S. A. 980, 991, 1068, 1075  
 Kumar, V. 1086, 1088  
 Kurata, M. 196, 233  
 Kurdjumov, G. V. 961, 977, 979, 991, 1065, 1075  
 Kurz, W. 645, 647, 657, 661, 682, 683, 699, 701
- La Mer, V. K. 625, 632, 700  
 Laaser, W. 610, 620  
 Lacombe, P. 845, 858  
 Laidler, K. J. 79, 94  
 Laird, C. 505, 527, 764, 795, 953, 958  
 Laird, J. G. 770, 795  
 Lally, J. S. 310, 325  
 Lambot, H. J. 834, 858  
 Landau, L. D. 214, 233, 827, 831, 1064  
 Lang-Weise, A. 801, 806, 817  
 Lange, W. F. 456, 479  
 Lange, W. F. III 715, 717, 795  
 Langer, J. S. 504, 527, 557, 620, 660, 661, 700, 729, 751, 795  
 Larsson, L. E. 769, 793  
 Laughlin, D. E. 830, 831, 860, 877, 885, 886, 893, 895, 896, 957 8  
 Laves, F. 121, 169, 782, 794, 885, 908, 958  
 Lavrentyev, F. F. 900, 960

- Lawley, A. 145, 169  
 Lawson, A. W. 207, 233  
 Lay, S. 948, 958  
 Lazarus, D. 143 5, 167-8, 382, 420  
 Le Claire, A. D. 93 4, 396 7, 404 5, 408-9, 412, 415, 420  
 Le, T. D. 891, 960  
 Leamy, H. J. 562, 563, 620  
 Lee, C. 144-5, 170  
 Lee, J. K. 449, 456, 479  
 Lee, K. 558, 620  
 Leibfried, G. 285, 298, 325  
 Lelie, H. M. M. V. D. 1099, 1101  
 Lement, B. S. 749, 795  
 Lennard-Jones, J. E. 117, 164, 169, 199, 354, 423  
 Leslie, W. C. 746-8, 795, 847, 858, 891, 960  
 Levine, H. 659, 662, 700  
 Levine, M. M. 584 6, 619  
 Levine, P. 918, 958  
 Lewis, B. G. 1080, 1088  
 Lewis, B. 604, 620  
 Lewthwaite, G. W. 255, 325  
 Li, C. H. 376 7, 836, 858  
 Li, J. C. M. 344 5, 361, 377  
 Li, Y. Y. 224, 228, 233 4, 1067, 1075  
 Lidiard, A. B. 103, 105, 382, 396 7, 404 6, 420  
 Lie, K. H. 142 3, 167  
 Lieberman, D. S. 6, 22, 995, 1004, 1060, 1105, 1112  
 Liebmann, B. 847, 858  
 Lifshitz, E. M. 214, 233, 827, 831  
 Lifshitz, F. M. 785, 795  
 Lighthill, M. J. 568, 620  
 Lin, J. 430, 479  
 Linde, J. O. 228, 233  
 Lipson, H. 125, 168, 721, 793, 821, 822, 831  
 Lipton, J. 661, 700  
 Liu, M. 1066, 1075  
 Liu, Y. C. 304, 324, 518, 522, 527, 956, 958, 1042, 1060  
 Livingston, J. D. 643, 700, 787, 795  
 Lloyd, L. T. 901, 906, 958  
 Loberg, B. 374, 377  
 Lodge, K. W. 356, 377  
 Logan, R. A. 759, 795  
 Lomer, W. M. 112, 118, 128, 130, 133, 138, 169, 272, 301 2, 305, 325, 338, 377, 418 20, 994, 1047, 1060  
 Longuet-Higgins, H. C. 198 9, 233  
 Loper, C. R. 628, 629, 700  
 Loretto, M. H. 305, 307, 313, 323 5  
 Lothe, J. 267 9, 285, 324 5, 424 6, 430 1, 440 3, 451, 479  
 Lothe, J. 577, 620, 627, 700, 938, 940, 958  
 Louat, N. 273, 325  
 Low, J. R. 253, 280, 285, 325  
 Lu, S.-Z. 646, 662, 683, 699-700  
 Luborsky, F. E. 1080, 1088  
 Lücke, K. 744, 795, 847, 849, 858  
 Lukens, H. C. 21-2  
 Lux, B. 687, 700  
 Lyons, J. V. 695, 700  
 Machlin, E. S. 483, 527, 987, 991, 1020, 1023, 1025, 1042, 1060, 1065, 1075  
 Mackay, A. C. 617-8, 620  
 Mackenzie, I. K. 136, 169  
 Mackenzie, J. K. 25, 78, 965, 990, 993, 1029, 1031, 1034, 1036-8, 1060  
 Mackenzie, K. K. 1018-20, 1022, 1059  
 Maclean, D. 316, 323, 341, 376  
 Maddin, R. 141-2, 169  
 Madonna, L. A. 559, 620  
 Magee, C. L. 885, 953, 958  
 Mahajan, S. 890 1, 900 1, 909, 921, 928, 944 5, 949, 951, 954, 956 60  
 Maitrepierre, P. L. 1080, 1088  
 Makino, A. 1080, 1088  
 Malen, K. 267, 325  
 Malin, A. S. 1047, 1059  
 Malis, T. F. 354, 377  
 Mallejac, D. 135, 167, 169  
 Mann, E. 133, 170  
 Manning, J. R. 397, 405 7, 420  
 Mar, R. W. 573, 620  
 Maradudin, A. A. 277, 325  
 Marburger, R. E. 1072, 1075  
 Marcinkowski, M. J. 821, 831  
 Marder, A. R. 810, 817  
 Margolin, H. 1042, 1060  
 Markovitz, H. 541, 552  
 Marquese, J. A. 786 7, 795  
 Marth, P. E. 456, 479  
 Martius, H. R. 651, 699  
 Maruyama, M. 564, 620  
 Mason, J. T. 683, 701  
 Mason, K. L. 533, 552  
 Massalski, T. B. 107, 167, 709, 717  
 Masters, B. C. 310, 324 5  
 Masumoto, T. 1080, 1087  
 Matano, C. 390-1, 403, 420  
 Matas, S. J. 1089, 1101  
 Mathewson, C. H. 908, 958  
 Mathot, V. 199, 234  
 Matthews, J. W. 598, 621  
 Maximova, O. P. 961, 977, 991  
 Mayer, J. E. 131, 161, 169, 277  
 Mazey, D. J. 403, 420  
 McDonald, A. B. 136, 169

- McDougall, P. G. 759, 796, 1021, 1060  
 McHargue, C. J. 891, 960  
 McKee, T. B. A. 136, 169  
 Meakin, J. D. 145, 169  
 Mehl, R. F. 18, 22, 510, 512, 527, 530, 536-7, 552, 805, 810, 813-4, 816 7, 838-9, 857-8, 977, 990  
 Mehrer, H. 136 7, 143 5, 170, 397, 412, 420  
 Meijering, J. L. 121, 163, 169, 334, 377  
 Meiron, D. 659, 700  
 Meisel, L. V. 545, 552  
 Mendenhall, C. E. 624, 700  
 Menguy, N. 618, 621  
 Meshii, M. 142, 144, 169  
 Meyer, J. 610, 620-1  
 Mikkola, D. E. 893 4, 958  
 Milgram, M. 131, 133, 168  
 Millard, D. J. 913, 916, 932, 935, 948, 958  
 Miller, C. E. 564, 620  
 Miller, M. K. 736, 750, 795  
 Miller, W. R. 631, 700  
 Miltat, J. E. A. 250, 325  
 Minonishi, V. 925, 927, 958  
 Miodownik, A. P. 110, 169  
 Mirzaev, D. A. 710, 712, 713, 717  
 Misra, R. D. 116, 169  
 Mitchell, J. W. 316, 324  
 Mitchell, T. E. 913, 959  
 Miura, S. 143, 170, 944, 958, 1107, 1112  
 Miyazaki, T. 630, 789, 791 2, 794 5  
 Miyazawa, Y. 627, 700  
 Modin, S. 805, 817  
 Molinder, G. 813, 815 7  
 Monroe, E. 161, 169  
 Mooradian, V. G. 979, 990  
 Moore, A. J. W. 910, 958  
 Moore, K. I. 633 4, 700  
 Morgan, E. R. 989, 991  
 Mori, T. 142, 144, 169, 890, 913, 944 6, 957 8  
 Morozov, O. P. 703, 717  
 Morris, L. R. 675 6, 679, 700  
 Morrogh, H. 687, 700  
 Morton, A. J. 305, 325  
 Moss, S. C. 211, 214, 233, 828, 831  
 Mott, N. F. 117, 129, 164 5, 169, 202, 233, 243, 267, 325, 344 5, 377, 485, 527, 849, 858  
 Muddle, B. C. 1024, 1028, 1060 1  
 Mughrabi, H. 315, 325  
 Mukhopadhyay, N. K. 619, 621, 1085, 1088  
 Muldawer, L. 227, 233  
 Mullen, J. G. 93 4, 397, 409, 420  
 Mullins, W. W. 295, 495, 497, 527, 569, 621, 660, 676, 700  
 Munster, A. 229, 234  
 Murakami, Y. 1107-10, 1112  
 Murphy, W. K. 228, 234  
 Musso, E. 1080, 1088  
 Mutaftschiev, B. 558, 619  
 Müller, E. W. 146, 169  
 Müller-Krumbhaar, H. 504, 527, 660, 661, 700  
 Nabarro, F. R. N. 122, 169, 202, 233, 243, 256-7, 266-7, 273-8, 280-1, 284 5, 301, 325, 358, 465, 468 70, 479, 611, 621, 933, 958  
 Nachtrieb, N. H. 94, 128, 131, 169, 380, 420  
 Nadeau, J. 655, 701  
 Nagakura, S. 750, 794-5  
 Nakanishi, N. 1035, 1060, 1105, 1112 3  
 Nakatani, I. 614, 621  
 Nakkalil, R. 816 7  
 Narita, N. 890, 913, 944, 958  
 Nash, C. E. 659, 700  
 Nehrenberg, A. E. 815, 817  
 Nesbitt, L. A. 896, 958  
 Neumann, P. 42, 78  
 Newkirk, J. B. 227, 234, 708, 717, 820, 831  
 Newman, R. C. 266, 323, 507, 527, 544 6, 552  
 Nicholas, J. F. 40 2, 78, 295 7, 301, 304, 324  
 Nicholson, M. E. 226, 234, 805, 817  
 Nicholson, R. B. 776 7, 779 81, 787, 793, 795  
 Nielsen, A. 578, 580, 621  
 Nihoul, J. 146, 169  
 Nishioka, K. 426, 431, 442, 479  
 Nishiyama, Z. 1053, 1060  
 Niwa, K. 199, 234  
 Nix, W. D. 280, 324  
 Norden, H. 374, 377  
 Northcott, L. 689, 700  
 Nouet, G. 948, 958  
 Nowick, A. S. 135, 168 9, 181, 189, 198, 206, 233, 823, 826, 831  
 Nutting, J. 749, 776 7, 779, 781, 784, 794 5, 1052 3, 1057, 1060  
 Nye, J. F. 78, 104 5, 316, 320, 325, 338, 377, 976, 991, 1048, 1060  
 Nyström, J. 767, 769, 795  
 Oblak, J. C. 1092, 1101  
 Oblak, J. M. 893 4, 958  
 Obretenov, W. 579, 621  
 Odinokova, L. P. 900, 960  
 Oehring, M. 1082, 1088  
 Ogawa, K. 947, 958  
 Ohara, M. 586, 621  
 Ohm, R. K. 1099, 1101  
 Ohta, Y. 2, 22  
 Oka, M. 1099, 1101

- Okamoto, P. R. 42, 78  
 Okamoto, H. 1099, 1101  
 O'Keefe, M. 409, 420  
 Olander, A. 1102, 1113  
 Oliver, D. S. 933, 958  
 Olson, G. B. 15, 21 2, 919, 958, 980, 991, 1049, 1060, 1064, 1067 8, 1075  
 Onsager, L. 95-105, 153-5, 169, 191, 224, 234, 392-3, 407, 420  
 Ookawa, A. 944, 958  
 Oriani, R. A. 198 9, 201, 206, 228, 234, 853, 858  
 Orlov, S. S. 267, 324  
 Orowan, E. 927, 934, 958  
 Otsuka, K. 972, 981, 984, 991, 1039, 1060, 1104, 1107, 1113  
 Otte, H. M. 42, 78, 965, 991, 999, 1022, 1025, 1039, 1060 1  
 Owen, W. S. 710 1, 717, 1062, 1075
- Paik, J. S. 630, 700  
 Paneth, H. R. 128, 169, 380, 420  
 Panseri, C. 783, 795  
 Papapetrou, A. 501, 527, 657, 700  
 Parcel, R. W. 810, 817  
 Parker, E. R. 376 7, 836, 858  
 Parr, J. G. 702, 714, 716 7  
 Parthasarathi, M. N. 847, 851, 858  
 Partridge, P. G. 310, 325, 900, 960  
 Pashley, D. W. 228, 233 4, 598, 621, 895, 958  
 Patel, J. R. 981 2, 991, 1064, 1075  
 Paton, N. E. 900, 960  
 Patterson, R. L. 1056, 1060  
 Paxton, A. T. 893, 958  
 Paxton, H. W. 891, 908, 958  
 Peach, M. O. 254 5, 260, 325  
 Pearson, D. D. 810, 817  
 Pearson, S. 141, 167  
 Pearson, W. B. 120, 169  
 Peierls, R. 273 4, 276 81, 284, 299, 308, 325, 358  
 Peissker, E. 913, 958  
 Penrose, R. 617, 621  
 Perepezko, J. H. 629 30, 700  
 Perini, A. 558, 621  
 Perrin, R. C. 120, 170, 278, 311, 323, 325  
 Petch, N. J. 1012, 1014, 1060  
 Petermann, J. 771, 795, 816 7  
 Peterson, N. L. 382 3, 412, 420  
 Pettifor, D. G. 2, 22, 113, 169  
 Pfaff, F. 280, 319, 325  
 Pfann, W. G. 671, 700  
 Philibert, J. 989 91  
 Philip, J. R. 494, 528  
 Philofsky, E. M. 400, 420  
 Pickering, F. B. 1089, 1092, 1101
- Piller, J. 1085, 1088  
 Piper, W. W. 595, 621  
 Pirouz, P. 901, 950-1, 958, 960  
 Pitsch, W. 744, 746-7, 795, 1052, 1060  
 Plaskett, T. G. 680, 700  
 Plichta, M. C. 714, 717  
 Polich, S. J. 595, 621  
 Polk, D. E. 1080, 1088  
 Pond, R. C. 354, 374 5, 377, 861, 890, 913, 923, 958 9  
 Pontikis, V. 563, 621  
 Portier, R. 375, 377  
 Potter, E. V. 21 2  
 Pound, G. M. 185, 234, 424 6, 430-1, 440 3, 451, 477, 479, 493 4, 501 2, 528, 557, 559, 572, 577, 598, 620 1, 625, 627, 630, 632, 700  
 Pound, M. 605, 621  
 Powell, E. G. 629, 700  
 Powell, H. M. 227, 233  
 Pratt, P. L. 910, 912, 956, 959  
 Presland, A. E. B. 228, 234  
 Preston, G. D. 774, 795  
 Price, P. B. 914, 959  
 Priestner, R. 891, 960  
 Prigogine, I. 104, 199, 234  
 Probstein, R. F. 451, 479  
 Protsenko, L. N. 899 900, 959  
 Pugh, S. F. 910, 959  
 Puls, M. P. 811, 817  
 Pumphrey, P. H. 354, 377  
 Pumphrey, W. I. 695, 700  
 Purdy, G. R. 504, 528, 752, 795
- Quimby, N. L. 826, 831  
 Quimby, S. L. 826, 831
- Rabier, J. 901, 960  
 Raghavan, V. 1062, 1071, 1074 5  
 Rahman, A. 163, 169  
 Ralph, B. 351, 376  
 Randlett, M. R. 110 11, 168  
 Ranganathan, S. 346 7, 349, 351, 376 7, 619, 621, 1085, 1088  
 Rao, B. V. N. 1029, 1060  
 Rapperport, E. J. 899, 903, 959  
 Rasmussen, D. H. 628 9, 700  
 Rassow, E. 908, 957  
 Rath, B. B. 847, 858  
 Rathenau, G. W. 803, 817, 845, 858  
 Rawlings, R. D. 280, 324  
 Ray, I. L. F. 299, 323  
 Ray, R. 1080, 1087 8  
 Raynor, G. V. 107, 169, 172, 233, 704, 717  
 Read, T. A. 8, 22, 931, 956, 965, 990 1, 1022, 1025, 1060, 1112

- Read, W. T. 247 52, 261, 267-8, 324-5, 336, 339 40,  
351, 356-60, 377, 463, 472, 581, 621, 700, 982,  
984, 1102
- Rechtien, J. J. 907, 959
- Redman, J. K. 890, 956
- Reed-Hill, R. E. 881, 898 900, 909, 959 60
- Rees, G. J. 604, 620
- Reichelt, K. 608, 621
- Reid, C. A. 912, 959
- Reid, C. N. 912, 959
- Reid, R. C. 586, 621
- Reiss, H. 423, 430 1, 439, 479, 558, 621
- Ren, X. 1107, 1113
- Resnick, R. 403, 420
- Reynolds, J. E. 1102, 1113
- Reynolds, W. T. 755, 795
- Rhines, E. N. 820, 831
- Rhines, F. N. 227, 234
- Rhodin, T. N. 599, 622
- Rice, S. A. 81, 94, 420
- Richards, C. E. 912, 959
- Richards, M. J. 211 14, 234, 886, 895, 959
- Richman, R. H. 810, 817, 889, 891, 953, 956,  
959 60
- Ridley, N. 810 1, 817
- Rieu, J. V. 893 4, 957
- Roberts, B. W. 227, 234
- Roberts, C. S. 898, 900, 957
- Roberts, G. A. 813 4, 817
- Robertson, D. 605, 621
- Robertson, I. M. 890, 960
- Robertson, W. D. 898, 959
- Robins, J. L. 608, 621 2
- Robinson, V. N. E. 608, 621
- Rodebush, W. H. 424, 479
- Roitburd, A. L. 1064, 1067 8, 1075
- Rosenbaum, H. S. 761, 763, 795, 910, 916, 959
- Rosenberg, A. 664, 690, 700
- Rosenfield, A. R. 891, 960
- Ross, J. 786 7, 795
- Ross, N. H. D. 1039, 1060
- Rothman, S. J. 412, 420
- Rothstein, J. 822, 826, 831
- Rowe, A. H. 1047, 1059
- Rowlands, P. C. 860, 891, 956, 959
- Rudman, P. S. 110, 165, 170
- Ruf, P. K. 456, 479
- Ruhl, R. C. 1080, 1088
- Rundman, K. B. 723, 732 4, 795
- Runnels, L. K. 162, 170
- Rushbrooke, G. S. 221, 234
- Russell, K. C. 426, 430, 440 1, 450 1, 456, 478 9
- Rutter, J. W. 350, 376, 671-2, 672, 674, 700, 847 9,  
856 8
- Saada, G. 124, 170, 300, 325
- Saburi, T. 1107, 1110, 1113
- Säfsten, K. M. 767, 793
- Sagel, K. 229, 234
- Saito, F. 659, 700
- Saka, H. 697, 700
- Sakamoto, H. 1105, 1113
- Salita, O. P. 900, 960
- Salzbrenner, R. J. 985-6, 991
- Sandvik, B. J. C. 1028 9, 1093, 1099, 1101
- Sandvik, B. P. 1029, 1060
- Sargent, G. M. 886, 956
- Sarian, S. 786, 795
- Sarrazin, T. 896, 959
- Sastry, G. V. S. 617, 621
- Sato, H. 1060
- Saxl, I. 271, 292, 325, 954, 959
- Scattergood, R. O. 261, 267 8, 323
- Schaefer, R. J. 657 8, 660, 699
- Schaeffer, P. C. R. J. 628, 700
- Schapink, F. W. 305, 325
- Schechtman, P. J. 165, 170
- Scheil, E. 629, 666, 684, 699 700, 739, 795, 801,  
806, 817
- Schey, J. 563, 621
- Schilling, W. 147, 170
- Schmid, E. 971, 973, 991
- Schober, T. 374, 376 7
- Schoeck, C. 654, 700
- Schoeck, G. 271 2, 280, 298 9, 323, 325
- Schottky, G. 411, 420
- Schroeder, T. A. 981, 991, 1107, 1113
- Schubert, K. 228, 234
- Schultz, H. 145, 170
- Schumacher, D. 146, 170, 411, 420
- Schwartzkopff, K. 310, 325
- Scott, G. D. 163, 170
- Scripou, V. 630, 700
- Searcy, A. W. 573, 620
- Sears, G. W. 505, 527, 612, 619, 621
- Seeger, A. 130 1, 133, 136 7, 143 5, 147, 170, 253,  
266, 272, 280, 298 9, 305, 309, 325, 382 3, 411 13,  
420, 938, 959
- Segall, R. L. 40 1, 78, 305, 307, 323, 325
- Seidman, D. N. 142 5, 167, 170, 308, 325, 411,  
421
- Seigle, L. L. 403, 420
- Seith, W. 770, 795
- Seitz, F. 108, 120, 122, 141, 169 70, 244, 251 3, 322,  
325, 356, 372, 379, 382, 394, 400, 402 3, 420, 543,  
782, 795
- Sekerka, R. F. 495, 497, 527, 660, 676-8, 700
- Selke, H. 618, 621
- Serra, A. 924-7, 949 950, 956, 959-60, 1058, 1060

- Seymour, W. E. 839, 858  
 Shaefer, R. J. 504, 527  
 Sham, L. J. 2, 22  
 Shapiro, J. M. 523, 528  
 Sharaf, M. A. 559, 621  
 Sharma, R. K. 144-5, 170  
 Sharma, S. M. 1082, 1088  
 Sharp, R. S. 670, 679, 700  
 Shawyer, R. 426, 431, 442, 479  
 Shechtman, D. 617, 621, 895, 959  
 Sheng, H. W. 698, 700  
 Shepard, G. 618, 620  
 Sherrill, F. A. 314, 326  
 Shewmon, P. 696, 701  
 Shewmon, P. G. 382, 420, 496-8, 528  
 Shiflet, G. J. 800, 817  
 Shih, C. A. 977, 991  
 Shimizu, K. 981, 1053, 1060, 1099, 1101, 1107, 1113  
 Shimoji, M. 199, 234  
 Shingu, P. H. 143, 170  
 Shiraiishi, S. Y. 629, 701  
 Shirn, G. A. 129, 138, 169, 381, 421  
 Shockley, W. 247, 261, 267, 279 80, 296, 298, 300,  
 302, 305 6, 308, 310, 324 5, 336, 351, 356 60, 377,  
 404  
 Shubnikov, A. V. 375, 377  
 Shuttleworth, R. 329, 377, 423, 478  
 Siegel, R. W. 142 5, 167, 170, 411, 421  
 Sikka, S. K. 1082, 1088  
 Silcock, J. M. 778 9, 781, 795  
 Silcox, J. 142, 169 70, 307, 325  
 Simmons, R. O. 135, 170, 412, 421  
 Simon, F. E. 166, 170  
 Simon, J. P. 356, 377  
 Sinclair, J. E. 2, 22, 113, 168  
 Sindzingre, P. 563, 621  
 Singh, A. 619, 621  
 Sinha, A. K. 120, 170  
 Sivertsen, J. M. 226, 234  
 Skylstad, K. 267, 324  
 Sleswyk, A. W. 311, 325, 921, 928 30, 937, 942, 947,  
 959  
 Slippery, W. A. Jr. 900, 959  
 Slyozov, J. 785, 795  
 Smallman, R. E. 142, 169-70, 300, 305, 308, 310, 324,  
 326  
 Smialowski, M. 673, 701  
 Smigelskas, A. D. 382, 421  
 Smith, C. S. 227, 233, 330, 333 4, 377, 633, 701, 740,  
 795, 853, 858  
 Smith, D. A. 127, 170, 354, 377  
 Smith, D. W. 805, 817  
 Smith, E. 245, 317, 322 3, 366, 376, 899, 959  
 Smith, F. W. 163, 170  
 Smith, G. V. 805, 817  
 Smith, J. H. 862, 900, 959  
 Smith, R. 1046, 1059  
 Smith, R. W. 704, 717  
 Smith, T. B. 856, 988, 990  
 Smith, V. G. 669, 701  
 Smoluchowski, R. 232, 234, 344, 377, 385, 411  
 Soffa, W. A. 830, 831  
 Sokolnikoff, I. S. 78  
 Sokurskii, Yu. N. 899-900, 959  
 Sommer, F. 1080, 1088  
 Southin, R. T. 633, 642, 701  
 Spaepen, F. 354, 376  
 Speich, G. R. 1093, 1101, 522, 527-8  
 Spencer, D. B. 693, 701  
 Sperry, P. R. 841, 857  
 Spiller, G. D. T. 610, 621  
 Spriet, B. 907, 959  
 Sree Harsha, K. S. 927, 959  
 Srinivasan, C. R. 1093, 1099, 1101  
 Stanley, J. K. 839, 858  
 Stanley, J. T. 508, 527  
 Starke, E. A. 893 4, 957  
 Staveley, L. A. K. 628, 701  
 Steeds, J. W. 305, 325, 942, 959  
 Steeple, H. 822, 831  
 Stehle, H. 272, 325  
 Stein, B. J. 610, 621  
 Stein, D. F. 280, 285, 325  
 Steinhardt, P. J. 918, 958  
 Stepakoff, G. 110, 170  
 Stewart, G. W. 162, 170  
 Stobbs, W. M. 299, 325  
 Stoloff, N. S. 896, 959  
 Stowell, M. J. 268, 324, 599, 604 6, 608, 621, 629 30,  
 701  
 Stranski, I. N. 561, 621  
 Straumanis, M. 684, 701  
 Stroh, A. N. 261 2, 267, 302, 325, 927, 957  
 Struden, J. L. 893 4, 957  
 Strunk, H. 592, 619  
 Suezawa, M. 1068, 1075  
 Sullivan, G. A. 135, 170  
 Sumino, K. 932, 959  
 Sun, Y. 772, 795  
 Sunagawa, I. 592, 621  
 Sundquist, B. E. 523 8  
 Surek, T. 573, 621  
 Suryanarayana, C. 1077, 1087 8  
 Sutton, A. P. 2, 22, 113, 170, 354, 377, 554, 556, 621,  
 918, 924, 959, 1058, 1060  
 Suzuki, H. 272, 278, 311, 325, 890, 908, 944, 951,  
 959  
 Swallow, E. 1093, 1101

- Swan, E. 184, 234  
 Swanger, L. A. 267, 323  
 Swann, P. R. 1035, 1055, 1061, 124, 168, 323  
 Swendsen, R. H. 586, 621  
 Sworn, C. H. 299, 325, 652, 701  
 Sykes, C. 219, 234, 820, 831  
 Szirmae, A. 839, 857
- Tadaki, T. 1107, 1113  
 Takagi, M. 628, 697, 701  
 Takagi, Y. 224, 234  
 Takahashi, T. 629, 632, 701  
 Takamura, J. 142 3, 170, 890, 913, 944, 958  
 Takeuchi, S. 965, 991  
 Tanaka, Y. 1030, 1060, 1099, 1101  
 Tanner, B. K. 652, 701  
 Tanner, L. E. 1080, 1087  
 Tarshis, L. A. 639, 658, 700 1  
 Tauer, K. J. 109 10, 170  
 Tawancy, H. M. 896, 959  
 Taylor, C. A. 821 2, 831  
 Taylor, G. I. 236, 325  
 Taylor, R. 119 20, 126, 167 8, 278, 311, 313, 323  
 Teghtsoonian, E. 653, 696, 701  
 Temkin, D. E. 502, 528, 658, 701  
 Tennevin, J. 834, 857  
 Terasaki, F. 891, 960  
 Tersoff, O. 2, 22  
 Tewordt, L. 130, 133, 170  
 Thalal, A. 375, 377, 918, 957  
 Thall, B. M. 687, 701  
 Thinel, J. 1102, 1112  
 Thomas, D. G. 628, 701  
 Thomas, G. 42, 78, 783 4, 795, 1029, 1060  
 Thompson, N. 301, 303 7, 325, 911, 913, 916, 932, 935, 948, 957 8  
 Thomson, J. J. 182 4, 234, 427, 487, 490, 495 9, 501 4, 512, 516, 523, 526  
 Thomson, R. 277, 324  
 Thornton, P. H. 836, 858  
 Thornton, P. R. 913, 959  
 Tiedma, T. J. 768, 795, 845, 850, 857 8  
 Tiller, W. A. 185, 234, 502 3, 527, 629, 631 2, 654, 658, 664, 667, 669, 671, 679, 690, 698, 700 1  
 Titchener, A. L. 832, 858  
 Tokuyama, M. 786 7, 794 5  
 Tolman, R. C. 423, 479  
 Tomsett, D. 955, 959  
 Tomsett, D. I. 292, 325  
 Tong, H. C. 980, 991  
 Treaftis, H. C. 509, 528  
 Treaftis, H. N. 767, 769, 780, 784 5, 795
- Trivedi and Kurz 648  
 Trivedi, R. K. 755–7, 794  
 Trivedi, R. 183, 234, 493 4, 501–5, 527–8, 645, 647 8, 657–8, 661, 682–3, 701  
 Trofimov, I. 608, 621  
 Troiano, A. R. 961, 991 2, 1018 21, 1023, 1025, 1060, 1098, 1101  
 Truesdell, C. 102–3, 105, 420  
 Tsai, A. P. 1085, 1088  
 Tsang, W. T. 597, 621  
 Tsou, A. L. 747, 795  
 Tsuzuki, K. 1067, 1075  
 Tu, K. N. 509, 525, 528, 901, 960  
 Tufte, O. N. 705, 716  
 Tuominen, S. M. 456, 479  
 Turkalo, A. M. 253, 325  
 Turnbull, D. 162, 164 5, 170, 382, 413, 415 6, 420 1, 434, 441 5, 448 9, 455, 479, 484, 509, 517 18, 522 3, 525 8, 624 5, 627, 630 3, 638, 641, 698 9, 701, 719, 742, 761 3, 767, 769 70, 780, 782 5, 794 5, 807, 817, 849 50, 857 8, 901, 960, 979, 990, 1064 5, 1067, 1075  
 Turski, L. A. 557, 620  
 Tweet, A. G. 759 60, 795
- Uhlmann, D. 643, 701  
 Urquhart, A. R. 184, 234  
 Usher, B. F. 608, 622  
 Uyeda, R. 560, 611, 614 5, 622
- Vaidya, S. 900, 949, 959  
 van Beijeren, H. 562, 622  
 Van Bueren, H. G. 146, 170  
 Van den Beukel, A. 147, 170  
 van der Hoek, B. 585 6, 622  
 van der Merwe, J. H. 294, 324, 358, 363, 365, 377, 598, 622  
 van der Ven, A. 755, 795  
 van Leuwen, C. 579, 619  
 Vandermeer, R. A. 838, 858  
 Vandershaeve, G. 896, 901, 959  
 Varley, J. H. O. 122, 169  
 Vasudevan, P. 1092, 1101  
 Velfe, H. D. 608, 622  
 Venables, J. A. 599, 601, 603 5, 607, 610, 620, 622, 890, 909, 934, 937, 941 4, 953, 959, 1059, 1061  
 Verbraak, C. A. 1099, 1101  
 Verhoeven, J. D. 810, 817  
 Verma, A. R. 589, 591, 622  
 Vermilyea, D. A. 568, 619  
 Vincent, R. 603, 622

- Vineyard, G. H. 80, 91 4, 129, 131, 133, 168, 170, 408, 421  
 Viola, M. 580, 622  
 Vitek, V. 120, 125 7, 170, 277 8, 280, 311, 313, 323-6, 353 4, 377, 891, 912, 918 9, 921, 924, 932, 947, 956, 959-60  
 Vitásek, E. 277, 322  
 Vlachavas, D. S. 375, 377, 861, 923, 959  
 Vohringer, O. 952, 959  
 Vold, C. M. 631, 699  
 Volmer, M. 422, 428 9, 431 4, 436-9, 441, 453 4, 479, 559, 622  
 Von Jan, R. 133, 170  
 Vonnegut, B. 624, 630, 633, 701  
 Voorhees, P. W. 786 7, 795  
 Votava, E. 928, 930, 959  
  
 Wade, K. D. 456, 479  
 Wagner, C. 785, 796  
 Wainwright, T. E. 162, 167  
 Waitz, T. 965 8, 991, 1055, 1061, 1067, 1075  
 Wajda, E. S. 129, 138, 169, 381, 421  
 Wald, M. S. 351, 376  
 Walker, C. B. 227, 234, 780, 796  
 Walter, J. L. 854, 858  
 Walther, G. 403, 420  
 Walton, D. 599, 601, 622, 675, 690, 701  
 Wan, X. 662, 701  
 Wang, C. C. 633, 701, 144 5, 170, 411, 421  
 Wang, N. 618, 622  
 Wannier, G. H. 224, 234  
 Ward, M. S. 113, 169  
 Ward, R. G. 629, 701  
 Warlimont, H. 1035, 1053, 1055, 1061  
 Warren, B. E. 124, 170, 227, 233  
 Warrington, D. H. 324, 347, 374, 377, 835, 858  
 Washburn, J. 300, 325, 376 7, 655, 701, 836, 858  
 Wasilewski, R. J. 1102, 1113  
 Wassermann, G. 1102, 1112  
 Watanabe, M. 1021, 1023, 1061  
 Watari, T. 196, 233  
 Watson, J. D. 759, 796  
 Wayman, C. M. 6, 22, 972, 980 1, 984, 991, 1019, 1021 5, 1028, 1056, 1059 61, 1093, 1099, 1101, 1107 8, 1113,  
 Weaire, D. 113, 127, 168  
 Weart, H. W. 786, 795  
 Webb, W. W. 612, 622  
 Weber, A. 428, 436, 479  
 Wechsler, M. S. 993, 998 9, 1018, 1028, 1031, 1035, 1061  
 Weeks, J. D. 562, 622  
 Weertman, J. 255 6, 284, 326  
 Wei, C. T. 955, 960  
 Weinberg, F. 647, 663, 696, 701  
 Weinig, S. 1042, 1061  
 Weins, J. J. 787, 796  
 Weins, M. J. 352-4, 356, 377  
 Weissberg, L. R. 826, 831  
 Weiss, R. J. 109-10, 169-70, 218  
 Weissberg 823  
 Weizer, V. G. 118, 131, 168  
 Wells, C. 1089, 1101  
 Wenzl, H. 147, 170  
 Wert, C. 742 3, 745 6, 796  
 Wert, C. A. 80, 88 9, 92, 94, 408, 410, 421, 489 90, 495, 528, 539-44, 552  
 Wessel, E. T. 891, 960  
 West, D. R. 305, 324  
 West, G. W. 687, 699  
 Westlake, D. G. 289, 326, 899 900, 916, 959  
 Westmacott, K. H. 142, 169 70, 305, 326  
 Weymouth, J. W. 135, 170  
 Wheeler, J. A. 61, 78  
 Whelan, M. J. 299, 304, 323, 326, 783, 795, 834, 858, 960  
 Whipple, R. T. P. 415, 418, 421  
 White, C. L. 456, 479  
 Whitham, G. B. 568, 620  
 Wiedersich, H. 271 2, 323  
 Wilcox, W. R. 670, 701  
 Wilkens, M. 228, 234, 263, 326, 1035, 1061  
 Wilkinson, H. 219, 234  
 Williams, A. J. 1040, 1061  
 Williams, D. B. 767 8, 793, 796  
 Williams, D. F. 909, 958  
 Williams, E. J. 155, 167, 217 27, 233  
 Williams, J. M. 508, 527  
 Williams, S. 354, 376  
 Williams, W. D. 687, 700  
 Willis, J. R. 267, 323, 326  
 Wilsdorf, H. G. F. 142, 169  
 Wilson, A. J. C. 821, 831  
 Wilson, E. A. 703, 710, 712, 717  
 Wilson, E. J. 441, 478 9  
 Wilson, F. H. 345, 377, 846 7, 858  
 Winegard, W. C. 675 6, 679 80, 693, 699 700, 849, 852 3, 855, 857  
 Winterbottom, W. L. 573, 622  
 Wittels, M. C. 314, 326  
 Wolf, H. 299, 325  
 Wood, J. K. 276 7, 324  
 Wood, M. S. 763, 796  
 Wood, W. A. 836, 858  
 Woodilla, A. J. 989, 991  
 Woodruff, D. P. 561, 622  
 Wright, J. G. 616, 619

Wulff, G. 155–60, 170, 295, 441, 481  
Wynblatt, P. 786, 793

Yagi, K. 610, 622  
Yamaguchi, M. 278, 326, 896, 919, 921, 932, 960  
Yang, C. N. 224, 234  
Yatsuya, S. 614, 622  
Yeo, R. B. G. 978, 991  
Yoffe, E. H. 267, 326  
Yoo, M. H. 900, 947–8, 955, 960  
Yoshida, H. 778, 796  
Young, D. J. 103, 105, 392, 420  
Young, F. W. 314, 326, 573, 620  
Ytterhus, A. 144, 170

Zachariasen, W. H. 23, 78  
Zeldovich, J. B. 436–7, 439–40, 442, 449–50, 479,  
807, 817  
Zener, C. 80, 88–90, 92, 94, 108–10, 117, 129, 136,  
138, 170, 201, 206, 234, 313, 380, 408, 410, 421,  
488–95, 498–9, 501–5, 512–13, 515–17, 519, 521–2,  
528, 539–44, 552, 1035, 1061  
Zetts, J. S. 145, 168  
Zhang, D. L. 634, 697–8, 701  
Zhang, X. 634, 1067, 1075  
Zief, M. 670, 701  
Ziegler, G. 781, 796  
Ziman, J. M. 112, 170  
Zinnsmeister, G. 599–600, 622  
Zukas, E. G. 891, 960

# Subject Index

Alloy systems are indexed alphabetically, irrespective of major and minor components

- $\alpha$ -martensite 1058-9, 1092
- $\beta$  copper-aluminium alloys 1036
- $\beta$ - $\alpha$  transformation in uranium 993-4, 1047-8
- $\epsilon$ -martensite 1031, 1058-9
- $\epsilon$ -phase 1031
- $\epsilon$ Zmartensite 1014
- $\lambda$  interface 988-9
  
- Abnormal eutectic 684
- Absolute stability 660
- Adiabatic heating 1063
- Allotriomorphic growth 560, 751 3, 755, 765
- Allotropy 702
- Alloys
  - amorphous 1078 80, 1082
  - eutectic 684 5
  - hypereutectic 641
  - hypereutectoidal 797, 805, 812, 1089
  - hypoeutectoidal 797 8, 805, 1089
  - monotectic 634
- Aluminium
  - copper alloys 670, 692, 709, 714, 742, 774 83, 785, 802 3, 834, 946, 1031, 1034, 1038, 1055 6
  - copper nickel alloys 984 6, 1040, 1104 5
  - silicon alloys 641, 685 7, 742, 761 3
- Amorphous
  - alloys 1078 80, 1082
  - metals 554, 616, 1078
  - solids 611, 616 17, 1077-9, 1081 3
  - state 594, 599, 616 17, 631, 1078 85
- Angle
  - dihedral 630 1
  - Euler 1005 6
- Anisotropic surface free energy 560, 651, 659, 662 3, 683, 806
- Annealing twins 845, 854-6, 922, 924
- Anti-generating node 938
- Antiphase
  - boundaries 819, 917
  - domains 818-22, 1047
- Antisymmetry 861
- Arrays 645, 653, 656, 662, 675, 682, 686, 727, 749, 834, 836, 859, 899, 928-30, 946, 956, 968-9, 995, 1000-1, 1025, 1039, 1049-50, 1053, 1055, 1064-6, 1083
- Asymmetry 572, 611, 694, 756, 912, 1105
- Athermal martensite 977, 988, 990, 1069, 1072, 1074 formation 633, 977-8
- Atom probe 720, 736 7, 750
  - position-sensitive 1082
- Atomic mobility 599 600, 609, 631, 704, 729, 759, 761, 776, 819, 962, 1077, 1079, 1083, 1096
- Ausforming 1069
- Austenite 595, 686 7, 692, 710, 749 52, 755, 769, 797, 800, 802 9, 813 16, 979, 982, 989 90, 993, 1012, 1014 16, 1020, 1023, 1025 6, 1029 31, 1051, 1059, 1067, 1070, 1073, 1077 8, 1089 90, 1092 6, 1098 101, 1107
  - grain size 813, 815
- Austenitizing time 815
- Autocatalysis 987, 990, 1071
- Autocatalytic
  - formation 1074
  - nucleation 1069, 1100
- Avrami 707, 744, 749, 760, 838 9, 842
  
- Back stress effect 588
- Bagaryatski-orientation relation 805 6
- Bain strain 993, 1015 16, 1030
- Baimit 710, 712, 800, 962, 1029, 1089 90, 1093 101
  - granular 1095
  - inverse 1089 90
  - lower 1089, 1091 2, 1095 7, 1099 100
  - reaction 799, 1090, 1092, 1095 6
  - upper 1089 90, 1092, 1095 7, 1100
- Barrierless nucleation 698, 770, 1068
- Binary
  - equilibrium diagram 638
  - solid solution 734, 736

- Boundary  
 antiphase 819, 917  
 coherent 1049  
 curvature 767 8, 851 3  
 incoherent 709, 720, 740, 766 7, 771, 1050  
 semi-coherent 740, 845, 1049–50
- Bowles Mackenzie dilatation parameter 1006, 1023–4, 1033, 1099
- Bragg-Williams 564, 650, 823 4, 826 7
- Bullough–Bilby theory 993, 1000 1, 1003 4, 1006, 1012, 1015
- Bullough's model of twinning 901 2
- Burst 987 8, 990, 1069, 1074  
 phenomenon 987, 1073
- Butterfly martensite 1014, 1067
- Cadmium gold alloys 931, 979, 984, 993, 1032, 1036, 1102, 1104 5, 1107
- Capillarity 647 8, 657, 661, 683  
 approximation 557 8  
 length 648
- Carbon iron nickel alloys 888 9, 891, 977, 982, 993, 1014, 1018, 1029, 1055, 1057, 1097
- Cast iron 685 7  
 grey 686 7
- Casting 633, 665, 688 9, 692 4  
 chill 633 4
- Cell spacing 682
- Cellular 645, 655 6, 661 2, 673, 678 9, 690 1, 816  
 growth 662, 675, 679 80, 682 3, 690, 767, 769  
 precipitation 720, 816  
 structure 646 7, 673 6, 686, 767, 1092  
 to dendrite transition 646, 679 80
- Cementite 687, 742, 747, 749 51, 797, 803, 805 9, 811, 813, 815 16, 1089, 1092, 1100  
 pro-eutectoid 805 6
- Channels 694
- Chemical beam epitaxy (CBE) 594, 597
- Chill  
 casting 633 4  
 crystals 688 90  
 layer 689
- Cobalt  
 iron alloys 890, 956, 967, 1055, 1066 7, 1069  
 nickel alloys 966 7, 983, 1055
- Coefficient  
 condensation 607 8  
 diffusion 565, 603, 675, 721 2, 733, 745, 752 4, 761, 770, 782, 811, 1082  
 distribution 648, 665, 667, 669 70, 692  
 partition 680, 686  
 sticking 713
- Coherent  
 boundary 1049  
 interface 598, 698, 708, 711, 715, 751, 755, 758 9, 764, 845, 856, 914, 916, 919, 922, 924, 931, 933, 954, 956, 961, 964 5, 1049 52, 1054, 1057 8  
 precipitate 720, 724, 774, 787, 791, 954  
 spinodal 721, 726, 729 30, 733
- Cold work 704, 759, 761 2, 770, 783, 821, 832 3, 835, 838, 850, 854
- Colony 803 4, 809–10  
 eutectic 686
- Columnar  
 crystals 688 91, 693  
 zone 689 90, 693–4
- Combined orientation 860, 887, 895 6, 924
- Complementary twinning dislocation 921 2, 928 9
- Complete mixing in the liquid 665 6, 669 70
- Composition  
 eutectic 640, 685, 688, 692  
 eutectoidal 797, 799 800, 806, 813  
 stoichiometric 644, 886
- Compound twins 860, 880 1, 898, 915, 937, 956, 1010
- Computer simulation 554 5, 557, 562 3, 573, 578 9, 609, 639, 651, 727, 737, 756, 787, 919, 922, 925, 932, 947, 949, 1003, 1027, 1056 7
- Concentration  
 gradient 661, 677, 682, 746, 755  
 waves 819, 827
- Condensation  
 coefficients 607 8  
 complete 599, 605 8  
 extreme incomplete 606 8  
 gas 615 16  
 initially incomplete 606  
 rate of 611  
 vacancy 652, 654 5, 783  
 vapour 554, 566, 594, 614, 635, 1079
- Conditional spinodal 828
- Constitutional supercooling 648, 671 2, 674 8, 680 1, 686, 692 3
- Continuous  
 ordering 736, 818, 822, 828, 831  
 precipitation 718 20, 737, 742, 759, 765 6, 772, 774  
 reaction 759, 761, 779, 819, 821 2
- Convection 615 16, 647, 663, 665, 669, 671, 676, 692 4
- Copper  
 manganese alloys 978, 987, 1046  
 zinc alloys 709, 715 16, 767, 819, 831, 952, 1039, 1102, 1107

- Correspondence matrix 881 5, 888-9, 892, 896, 903, 936, 994, 1010, 1015, 1030, 1032, 1044
- Critical  
 disordering temperature 818  
 nucleation 559  
 nucleus 554, 559 60, 575-7, 598-600, 602, 606, 610 11, 626, 632, 637, 648, 656, 703, 713, 740, 782, 839-40, 935, 982, 1064-7
- Crystal growth  
 classical theory 553-9, 598 601, 623, 741, 840, 859, 881, 1065, 1067, 1071  
 dislocation theory 553, 834, 836, 993, 1026  
 island growth 553-4, 562, 578, 599 600, 604, 606 7, 609-10, 614  
 Stranski Krastanov 553, 599
- Crystal surface 561 2, 564, 569, 581, 587, 644, 834
- Crystallite 554, 560, 599, 614, 616, 627, 688
- Crystallization 631, 684, 704, 839, 1082 5  
 eutectic 1083  
 polymorphic 1083  
 primary 1083 4  
 reaction 1084
- Cubic hexagonal transformations 993, 1043, 1055
- Curvature  
 boundary 767 8, 851 3  
 correction 557 60  
 interface 631, 647, 657 8, 811
- Decagonal structure 1085
- Decanting experiments 664
- Decomposition  
 eutectoidal 797, 799, 802, 808  
 isothermal 730  
 spinodal 718 32, 736 7, 741, 772, 819, 828, 831, 1064  
 spontaneous 722, 915  
 thermal 595, 597
- Degree of order 818, 820 1, 825 6
- Dendrite 560, 645, 648 50, 653, 656 64, 672 3, 679, 682 4, 691 2  
 freely growing 656, 663, 682  
 growth 560, 644, 649, 653, 656 61, 663 4, 666, 672, 679 80, 682-3, 686, 690 2  
 isothermal 657  
 network 639  
 non-isothermal 659  
 structures 639, 644, 646 7, 673, 690  
 thermal 656, 661  
 tips 646 7, 657, 660 1, 663, 682 3, 690
- Dendrite arm 691, 693  
 primary 691 2  
 secondary 691 2
- Dendrite spacing 682-3  
 primary 692
- Dense random packing 1082
- Devitrification 1083
- Dichromatic colour 923
- Diffraction line 820
- Diffusion 556, 559, 566, 572, 588, 596, 598, 603 4, 613, 615, 639, 642 3, 647, 649-50, 654, 656-7, 659, 665 7, 669 70, 674, 679, 682, 685-6, 691, 720 5, 728-31, 736, 742, 745-6, 750, 753, 755-7, 759-64, 767, 769 72, 776, 782, 785, 792, 811, 813, 816, 818, 827, 850, 961 2, 966, 981, 989 90, 1039, 1047, 1064, 1077, 1095, 1098, 1100, 1107  
 coefficient 565, 603, 675, 721 2, 733, 745, 752 4, 761, 770, 782, 811, 1082  
 -controlled 596, 612, 657, 666, 668, 676, 746 7, 749, 752 3, 755, 759 60, 764, 808, 1097 8, 1100  
 field 572, 645, 657, 659, 677 8, 786  
 grain boundary 771 2, 816, 853  
 lattice 710, 742, 746, 783, 785, 837, 848 9  
 long-range 686, 818, 1047  
 point effect 612, 656  
 solid-state 665, 679, 683, 692  
 solute 645, 647, 649, 665, 667, 672, 674, 683, 723, 781  
 surface 564 6, 569, 577, 588, 596, 598, 603, 606, 608, 613  
 thermal 647 9, 657, 668
- Diffusional displacive transformations 757 8, 961
- Dihedral angles 630 1
- Directional solidification 645, 649, 661, 666, 683
- Disclinations 923
- Discontinuous  
 precipitation 719 20, 766 8, 771 2, 782, 804, 808  
 reaction 767, 772, 797, 813, 815 16
- Dislocation  
 density 573, 589 90, 652, 654 5, 746, 759 61, 834, 839 42, 844, 976, 1029, 1048 50, 1056, 1092, 1100
- DSC 923  
 emissary 928 31, 1067  
 mobility 654, 838, 927, 931, 949 50, 1053  
 pole 932 3, 936 9, 941 2, 949  
 supplementary displacement 923, 927, 949
- Disordered  
 interface 650 1, 698, 752, 755, 849  
 lattice 885, 917  
 phase 818 20, 1039  
 solid solution 703, 818, 828, 892  
 surface 563, 650 1
- Disordering reaction 819
- Dispirations 923
- Distribution coefficient 648, 665, 667, 669 70, 692
- Divorced  
 eutectic 685  
 pearlite 1096

- Double twinning 867, 880-1, 898-900, 904
  - dislocation 916
- Driving force 553, 579-80, 584-5, 596, 598, 635-7, 644, 647, 649, 651, 672, 692, 710, 712-13, 716, 719-20, 738-9, 742, 751, 753, 771-2, 815, 819, 833, 836, 840, 843, 846-8, 851-4, 914, 918, 929, 939, 979-84, 988, 1062-9, 1073-4, 1083-4, 1098, 1103, 1107-8
- Droplet technique
  - embedded 633-4
  - small 554, 624, 638, 641-2, 979, 1066
- DSC
  - dislocations 923
  - lattice 923-4
- Effect of impurities or inclusions 652, 654, 849, 852-3, 954, 1069
- Elasticity
  - rubber-like (or ferro-) 1102, 1105-7
- Electrical resistivity 708, 746, 761-3, 780, 782-4, 820, 826, 977
- Electron microprobe analysis 692
- Embedded droplet technique 633-4
- Embryos 557, 577, 600-1, 623, 637, 704, 713, 739, 741, 776, 807, 819, 934-5, 945, 947, 1041, 1064, 1066-7
- Emissary
  - dislocations 928-31, 1067
  - lattice dislocations 930
- Energy
  - stored 832-3, 835, 838, 840, 981
  - strain 584-5, 598, 696, 704, 710-11, 724-6, 736, 739-40, 785, 788, 833, 836, 839, 843, 854, 914, 934-5, 961, 966, 980-1, 984, 986, 1052, 1062-3, 1067, 1095, 1098
  - thermal 556, 650, 703, 710, 712, 837, 931
- Entropy 564-5, 602, 627, 726, 811, 823, 833, 849, 982, 1063
  - activation 849
  - of configuration 555-6
  - of fusion 684
  - of melting 626, 628, 650
  - of transformation 648
- Epitaxy
  - chemical beam 594, 597
  - molecular beam 594, 596-8
  - vapour phase 594
- Equi-axed 611, 644, 693, 711
  - crystals 689
  - grains 688
- Equilibrium
  - freezing point 623, 672
  - shape 560, 575-6, 696
- Euler angles 1005-6
- Eutectic
  - abnormal 684
  - alloys 684-5
  - colony 686
  - composition 640, 685, 688, 692
  - crystallization 1083
  - divorced 685
  - grain 684-6
  - interdendritic 692
  - microstructure 640, 684, 686
  - mixture 639-41, 684, 691, 1077
  - normal 684
  - point 641
  - reaction 633, 666, 692
  - simple 639-40
  - structure 640, 684-5
  - system 639-40, 684-5
  - temperature 639-42, 685, 688
- Eutectoidal
  - composition 797, 799-800, 806, 813
  - decomposition 797, 799, 802, 808
  - mixture 798
  - point 806
  - reaction 797-9
  - structure 813
  - temperature 805, 810, 1096
  - transformations 797-8
- Extrinsic twinning dislocation 918
- Faceting/faceting 685
- Ferrite 710-12, 750-3, 755, 797, 803, 805-9, 813-16, 962, 1012, 1078, 1089-90, 1092-6, 1098-100
  - pro-eutectoid 805, 1092, 1096
- Ferroelastic 1105
- Fibrous 611, 684-5, 687
- Field ion microscopy 720, 750, 777, 1082, 1085
- First-order phase change 554, 695, 818-19, 828
- Flakes 686-8
- Fluctuations 554, 557, 566-7, 613, 638, 650, 659-60, 670, 672, 677-8, 720-4, 726-9, 731-2, 734-8, 740-1, 823-5, 828, 839-40, 931, 934-5, 1064-8, 1085
- Frank-Bilby theory 1048-50
- Free volume 1079, 1082-3
- Freely growing dendrite 656, 663, 682
- Freezing 560, 623-4, 628, 631, 633-4, 639, 644-5, 648-51, 665-72, 674-5, 679, 684, 686-7, 689, 693-4, 704, 1077, 1083
- Gas condensation 615-16
- Generating nodes 932-3, 935-8, 942, 954

- Gibbs Thomson 647, 649, 657, 659, 681, 785
- Glass  
 -forming alloys 617, 1079 80  
 transition temperature 1077-8, 1082-3
- Glassy state 616, 1078
- Gleiter 614, 616
- Globular 687, 805
- Gradient  
 energy 722, 724 5, 734, 736, 741, 1067  
 thermal 653, 655, 661, 677 8
- Grain boundary 554 5, 560, 599, 631, 642, 675,  
 695 7, 708 11, 713 16, 719, 740, 742, 750 1, 753,  
 755, 759, 763, 765-8, 770, 803 4, 806 9, 813, 816,  
 819, 833 4, 836, 838, 841, 843-4, 846, 848 9,  
 852 3, 856, 859, 902, 918, 922, 931, 952, 967 8, 980,  
 986, 1048, 1064, 1066, 1068, 1091, 1096, 1098 9  
 diffusion 771 2, 816, 853  
 migration 764, 772, 841 3, 845 6, 848 9, 852 3, 856  
 steps for 922
- Granular bainite 1095
- Graphite 686 7, 900, 902
- Gravitational segregation 647
- Greninger Troiano orientation relation 1014
- Grey  
 cast iron 686 7  
 tin 704 7, 905
- Growth  
 allotriomorphic 560, 751 3, 755, 765  
 dendritic 560, 644, 649, 653, 656 61, 663 4, 666,  
 672, 679 80, 682 3, 686, 690 2  
 of amorphous solids 616 17  
 of nanocrystals 614 16  
 of perfect crystals 553 80  
 of quasi-crystals 617 19  
 of real crystals 580 93  
 of surface coating 593 610  
 of thin films 593 610  
 of whiskers 611 13  
 orientated 850  
 rate 553, 560, 564, 566, 568, 574, 578 80, 583,  
 586 90, 596 7, 600, 613, 623 5, 644 6, 648 9,  
 653 6, 659 60, 664 5, 670, 673 5, 678, 680, 682,  
 686 8, 691 2, 703 7, 710, 713, 719, 726, 728 9,  
 742, 744, 746 7, 752 3, 755 6, 771, 801, 803,  
 809 15, 839, 849 50, 853, 946, 978, 1083, 1097 8,  
 1100  
 subgrain 834, 836 7, 841  
 theories 553, 580, 645, 666, 682, 686 8, 745, 753,  
 810 11, 850  
 thermally activated 703, 705, 710, 712, 716, 914,  
 977, 1046 7  
 time-dependent 662
- Guinier-Preston (GP) zone 719 21, 741, 762, 772,  
 774, 776-9, 782
- Habit plane 616, 686, 715-16, 721, 747, 759, 880,  
 889, 891 2, 896, 898-9, 903-7, 914, 964-5, 967-70,  
 981 2, 987, 992, 995, 997 9, 1001 7, 1011 15,  
 1017-21, 1023 31, 1034 47, 1050-2, 1054, 1057,  
 1064, 1098-9, 1108, 1112
- Hall-Petch relation 952
- Heterogeneous  
 nucleation 560, 580, 608, 624, 627, 631 5, 639, 695,  
 715, 840, 979, 1066, 1069, 1084 5  
 nucleus 628, 689
- Hole nucleus 609
- Homogeneous  
 nucleation 558, 592, 609, 624, 627 33, 639,  
 694 5, 704, 761 2, 839 40, 927, 934 5, 1065 9,  
 1083 5  
 nucleus 632
- Homogenization 670, 761, 770
- Hypereutectic alloys 641
- Hypereutectoid alloys 797, 805, 812, 1089
- Hypoeutectoid alloys 797 8, 805, 1089
- Icosahedral structure 1085
- Impurity 561, 571 2, 613, 624, 633, 652 3, 655  
 content 624, 704, 706 7, 848 50  
 Inclusions 652, 654, 849, 852 3, 954, 1069
- Incoherent  
 boundary 709, 720, 740, 766 7, 771, 1050  
 interface 598, 703, 708, 710 11, 714 16, 752, 756,  
 813, 856  
 nucleation 599
- Incomplete reaction phenomenon 1095, 1100
- Indium-thallium alloys 931, 979, 984 5, 987 9, 993,  
 1045 6, 1051, 1053, 1102, 1105 7, 1111
- Induction period 707 8, 824 5, 838
- Inoculation 704 5, 762
- Instability condition 677
- Interdendritic  
 eutectic 692  
 liquid 644, 684, 692  
 region 646, 692, 694  
 spacing 682, 691
- Interface  
 coherent 598, 698, 708, 711, 715, 751, 755, 758 9,  
 764, 845, 856, 914, 916, 919, 922, 924, 931, 933,  
 954, 956, 961, 964 5, 1049 52, 1054, 1057 8  
 curvature 631, 647, 657 8, 811  
 disordered 650 1, 698, 752, 755, 849  
 incoherent 598, 703, 708, 710 11, 714 16, 752, 756,  
 813, 856  
 interphase 634, 715, 727, 757, 808, 1046, 1048  
 isothermal 649, 657 8, 661  
 semi-coherent 598, 759, 778, 914, 1000, 1051, 1064

- Interfacial free energy 598, 631, 637 8, 642, 648, 660, 683, 720, 739, 741, 747, 785, 788, 811, 833, 839, 855, 1065–8
- Interlamellar spacing 686, 766, 770, 772, 805, 810, 814
- Intermediate phase 644
- Intermetallic compound 595 6, 644, 650, 685, 702, 1085
- Interphase interface 634, 715, 727, 757, 808, 1046, 1048
- Intrinsic twinning dislocation 918
- Invariant plane strain 756, 764, 881, 962, 964 5, 970 1, 981, 993, 995, 1017, 1023, 1041, 1048, 1051, 1064, 1068, 1093
- Inverse
  - bainite 1089 90
  - melting 1079, 1081
  - segregation 694
  - transformation 705, 984, 1112
- Iron
  - manganese nickel alloys 977 8, 1071, 1074
  - nickel alloys 711, 716, 845, 891, 978 81, 983 4, 987, 1014, 1019, 1021 2, 1056, 1062, 1072, 1074, 1077 8
- Irrational twinning elements 881, 889, 898
- Ising model 557 8
- Island nucleus 609 10
- Isothermal
  - decomposition 730
  - dendrite 657
  - interface 649, 657 8, 661
  - martensite 977, 988, 990, 1069
  - martensite formation 977 8, 1071
  - nucleation 978, 1065
  - reaction 718, 730 1, 749, 752, 782, 797, 833, 977, 1069
  - transformation 703, 710, 730, 759, 761, 797 8, 800 3, 814, 820, 977, 990, 1071 2, 1090 1
  - transformation curves 625 6, 708, 726, 814, 977 8
- Isotropic surface free energy 575, 659, 662, 676, 1085
  
- Jump frequency 713
  
- Kinetics of recrystallization 838, 842
- Kurdjumov Sachs orientation relation 752, 805, 1005, 1014, 1020, 1024, 1030, 1098 9
- Lamellae 684 6, 688, 766, 781, 797, 803, 805, 813, 856, 908, 911, 930, 954, 1000, 1055, 1093
- Lamellar 684 5, 801, 803, 806, 815 16
  - structure 687, 720, 797, 813
- Landau expansion 1064
- Laser ablation 615
- Latent heat 564, 585, 623, 628, 639–40, 644, 648, 650, 658, 677, 683, 694, 696, 713, 1021, 1063
- Lath martensite 710, 750, 1029–30
- Lattice
  - correspondence 755 6, 758 9, 764, 888, 915, 961, 993, 999, 1010, 1044, 1047, 1064, 1098
  - diffusion 710, 742, 746, 783, 785, 837, 848 9
  - disordered 885, 917
  - DSC 923 4
  - instability 1064, 1067 8
  - invariant deformation 901 2, 963 4, 992, 995 1000, 1003 4, 1011 12, 1016 17, 1023 8, 1030, 1033, 1040 1, 1046, 1048, 1050 1, 1055, 1066, 1099, 1112
- Laue 612, 749
  - photographs 653
  - units 735
- Ledge theory 755 6, 758 9, 764, 778, 843 4, 846, 1096
- Lenticular martensite 969, 1014, 1067
- Liquid plus solid region 631, 633, 639
- Local equilibrium at the interface 676, 753, 1100
- Long-range
  - diffusion 686, 818, 1047
  - field 555, 598, 928
  - order 616 17, 694, 818 26, 828, 885, 893 4, 1107, 1112
  - strains 1040
  - stresses 975, 1048 51, 1055, 1064
- Lower bainite 1089, 1091 2, 1095 7, 1099 100
  
- m** restriction curve 1002
- Macrosegregation 694
- Manganese titanium alloys 1043
- Marginal stability 660 3
- Martensite 709 10, 712, 716, 747, 749 51, 759, 800, 805, 862, 888 92, 897, 901, 905, 914, 956, 962 5, 968 9, 974, 977 82, 988 90, 992 5, 1000, 1010, 1012 16, 1018 21, 1023, 1025 6, 1028 31, 1034, 1036, 1039 40, 1043, 1045, 1047, 1050 6, 1058, 1062 71, 1073 4, 1077, 1089 90, 1092 3, 1098 100, 1102 3, 1105, 1107 12
  - $\alpha$ - 1058 9, 1092
  - $\epsilon$ - 1031, 1058 9
  - $\epsilon Z$  1014
- athermal 633, 977 8, 988, 990, 1069, 1072, 1074
- butterfly 1014, 1067
- isothermal 977 8, 988, 990, 1069, 1071
- lath 710, 750, 1029 30
- lenticular 969, 1014, 1067
- thermoelastic 909, 972, 1069
- thin-plate 1014, 1021 2, 1024, 1054 5, 1067, 1099

- Martensitic**  
 crystallography 897, 901, 965, 992, 1021, 1044, 1047, 1050, 1093  
 nucleation 710, 716, 1059, 1062 3, 1065, 1068 9, 1077  
 reaction 712, 800, 802, 961  
 transformation 702, 709-10, 713, 716, 749, 764, 820, 859, 885, 893, 911, 914, 931, 936, 947, 961-5, 977-9, 982, 984, 1000, 1012, 1016, 1036, 1039, 1043, 1047, 1051, 1062, 1064, 1069, 1074, 1077, 1098
- Massive** 612, 686, 709, 711 12, 715-16, 1077, 1100  
 reaction 712, 716, 802  
 transformation 709-10, 712, 714-15, 1062, 1077 8
- Maximum velocity hypothesis** 656, 660
- Melt**  
 spinning 697 8, 716, 1076 7  
 undercooled 645, 647, 649
- Melting** 623, 628, 636, 644, 670, 693 8, 849, 1077  
 inverse 1079, 1081  
 point 563 4, 576, 623, 626 8, 630, 632 3, 636, 649 50, 652, 654, 694 8, 730, 819, 853, 979, 1076 7, 1079, 1081, 1083  
 surface 563 4, 650, 695, 1077  
 zone 670
- Metallic glasses** 616 17, 1082
- Metastable liquidus and solidus temperatures** 640, 642 3
- Microsegregation** 691 2, 1077
- Microstructure**  
 eutectic 640, 684, 686  
 non-equilibrium 643
- Mid-rib** 1014, 1021 3, 1056, 1099
- Misfit** 598, 616, 634, 653, 655, 708, 742, 759, 761, 772, 788 9, 792, 849, 919, 931, 992, 1024, 1030, 1052
- Mixture**  
 eutectic 639 41, 684, 691, 1077  
 eutectoidal 798
- Mobility**  
 atomic 599 600, 609, 631, 704, 729, 759, 761, 776, 819, 962, 1077, 1079, 1083, 1096  
 boundary 617, 703, 836-7, 840, 845 7, 849 50, 854, 856, 984, 1000, 1053, 1068, 1100, 1108  
 cluster 600, 602 3, 605-6, 608 9  
 dislocation 654, 838, 927, 931, 949 50, 1053
- Modification** 641, 687
- Molecular**  
 beam epitaxy (MBE) 594, 596 8  
 dynamics 555 6, 558  
 statics 555 6
- Molybdenum-titanium alloys** 1041, 1043 4
- Monotectic**  
 alloys 634  
 reaction 633, 643, 684  
 system 695
- Monte Carlo simulation** 556 8, 562, 580, 586, 651, 737
- Nanocrystalline materials** 554, 614, 616, 1082
- Nanoscience** 611
- Nanotechnology** 611
- Nearest-neighbour** 556, 562, 564 5, 637, 650, 736, 739 40, 792, 924, 950, 1082
- Nickel titanium** 894, 1102
- Nishiyama Wassermann orientation relation** 1014, 1030, 1098 9
- Nodules** 685, 687, 801, 803, 805 6, 809
- Non-equilibrium microstructures** 643
- Non-faceting/faceting** 685 6
- Non-faceting/non-faceting** 685 6, 688
- Normal**  
 eutectic 684  
 segregation 693
- Nucleation**  
 autocatalytic 1069, 1100  
 barrierless 698, 770, 1068  
 critical 559  
 heterogeneous 560, 580, 608, 624, 627, 631 5, 639, 695, 715, 840, 979, 1066, 1069, 1084 5  
 homogeneous 558, 592, 609, 624, 627 33, 639, 694 5, 704, 761 2, 839 40, 927, 934 5, 1065 9, 1083 5  
 incoherent 599  
 isothermal 978, 1065  
 martensitic 710, 716, 1059, 1062 3, 1065, 1068 9, 1077  
 mechanism of 729, 767, 839, 918  
 operational 838, 840, 847, 978, 1064, 1066, 1069  
 quasi-steady-state 602, 623, 630  
 rapid 573, 607, 623, 628, 632, 637 8, 689, 696, 810, 821, 1072, 1089  
 rate of 557 9, 573, 577 8, 598 603, 606 7, 610, 623, 625 6, 628, 630 3, 637 8, 640 1, 703 4, 707 8, 714 15, 743 4, 746, 751, 753, 801 3, 806 10, 812 15, 838 40, 978, 1065 6, 1068 70, 1083, 1085, 1096 7  
 spontaneous 931 2, 968, 1065 6, 1068  
 steady-state 573, 578 9, 598 9, 601, 606, 623, 626, 839, 1084  
 sympathetic 987, 1069, 1098  
 thermal 712, 742, 747, 914, 1068, 1072, 1077  
 transient 601, 605 6, 623, 626, 708, 807, 809, 815  
 twin 933, 947

- Nucleus 553-4, 560, 562, 574-80, 592, 598-600,  
 609-10, 623-8, 632, 635-9, 641-2, 644, 653, 656,  
 684-5, 689-90, 694, 703-5, 707-9, 713, 715, 738-45,  
 747, 751, 753, 759-60, 762-3, 765, 767, 769-70, 782,  
 799, 803-5, 807-9, 812, 814, 819, 838-40, 843, 850,  
 914, 931, 933-5, 938, 967, 977-9, 987, 1046,  
 1064-70, 1072, 1078, 1083-5, 1095, 1098, 1105  
 cell 767, 769-70  
 coherent 715  
 critical 554, 559-60, 575-7, 598-600, 602, 606,  
 610-11, 626, 632, 637, 648, 656, 703, 713, 740,  
 782, 839-40, 935, 982, 1064-7  
 heterogeneous 628, 689  
 hole 609  
 homogeneous 632  
 incoherent 715  
 island 609-10  
 orientated 850-1, 880  
 semi-coherent 740  
 solid 623, 625, 635, 637-8, 640, 644, 656  
 stable 600, 605, 623, 625, 738-9  
 subcritical 600  
 supercritical 577, 632, 935  
 twin 880, 912, 914, 932-4
- Operations 867, 869, 881, 885, 888, 898, 938, 940-1,  
 1000, 1109  
 symmetry 861, 867, 884, 923, 999
- Order disorder  
 reaction 1046  
 temperature 818-9, 828  
 transformation 818-9, 828
- Ordered phase 751, 764, 818-20, 962, 1032, 1036,  
 1111
- Ordering reaction 736, 818-19, 821, 1046
- Orientated  
 growth theory 850  
 nucleus theory 850-1, 880
- Orientation relation 598-9, 634, 685-6, 697, 705, 714,  
 719, 751, 759, 778, 800, 805, 850, 859-61, 866-7,  
 875-6, 880-1, 888, 891, 897-9, 904, 907, 924,  
 997-8, 1004-5, 1014, 1020-1, 1024, 1029-31, 1034,  
 1036, 1038, 1040-3, 1047, 1054, 1098-9
- Bagaryatski- 805-6  
 Greninger Troiano 1014  
 Kurdjumov Sachs 752, 805, 1005, 1014, 1020,  
 1024, 1030, 1098-9  
 Nishiyama Wassermann 1014, 1030, 1098-9  
 Pitsch Petch 805-6
- Paraboloid of revolution 657-9  
 Paraequilibrium 753, 1094, 1100
- Partial  
 radial distribution function 1081  
 twinning dislocation 919-21, 924, 937
- Particle  
 irradiation 556, 704  
 ultrafine 611, 614-16
- Partition coefficients 680, 686
- Partitioning 645, 755, 785, 1070-2, 1085, 1092, 1094,  
 1098
- Pearlite 797, 800, 803-15, 1089-90, 1093, 1096,  
 1100  
 divorced 1096
- Péclet number 656-8, 661, 756, 758
- Penrose tiles 617-18, 1086
- Perfect interface dislocations 923
- Periodicity 618, 727, 918, 920, 1085  
 quasi- 618
- Peritectic  
 reaction 642-4, 666, 684  
 temperature 643
- Perturbations 660, 676-9, 828
- Phase  
 disordered 818-20, 1039  
 intermediate 644  
 ordered 751, 764, 818-20, 962, 1032, 1036, 1111
- Phase change 625, 704, 719, 832, 1062, 1105, 1109  
 first-order 554, 695, 818-19, 828  
 second-order 819, 828
- Pitsch Petch orientation relation 805-6
- Plain carbon steels 749, 800, 806, 811, 1073, 1096
- Planar  
 boundary 851, 853, 909, 915, 974  
 facets 698, 716  
 growth 648, 675-6, 683, 686, 690-1  
 interface 645-6, 648-50, 664, 666, 670, 672-8, 682,  
 686, 690-1, 698, 705, 715-16, 914, 919, 965, 984,  
 1048, 1054-5
- Plate-like 611, 692
- Plate-shaped dendrite 658, 692
- Platelet 613-15, 664, 749, 776, 781, 838, 914, 935  
 structure 567, 651, 664
- Point  
 eutectic 641  
 eutectoidal 806
- Pole  
 dislocation 932-3, 936-9, 941-2, 949  
 mechanism 933, 936-42, 944, 948-9, 954, 967-9  
 source 937, 939
- Polygonization 833-7, 840-1, 844, 846, 899
- Polymorphic  
 changes 702-4, 708, 710, 739, 818  
 crystallization 1083
- Polymorphism 702
- Position-sensitive atom probe 1082

- Precipitation  
 continuous 718–20, 737, 742, 759, 765–6, 772, 774  
 discontinuous 719–20, 766–8, 771–2, 782, 804, 808  
 Pre-exponential factor 559–60, 577, 579, 625–7, 632, 634–5, 1085
- Primary  
 crystallization 1083–4  
 dendrite arm 691–2  
 dendrite recrystallization 832, 837–8, 849, 851, 853–4  
 dendrite spacing 692  
 solid solution 644, 1043  
 Prismatic source 943–5  
 Pro-eutectoid 797–9  
 cementite 805–6  
 ferrite 805, 1092, 1096  
 Pseudo-elasticity 894, 1102–3  
 Pseudotwinning 880, 885, 893–5  
 Pure metal 595, 616–17, 623, 631, 633, 638–9, 644–5, 647–50, 656–7, 662, 664, 674, 681–2, 689–90, 693, 695, 702–3, 710, 739, 832, 875, 885, 912, 1079
- Quasi-crystal 611, 617–18, 918, 1085–6  
 one-dimensional 918  
 three-dimensional 618, 918  
 two-dimensional 617  
 Quasi-periodicity 618
- Radial distribution function 617, 1080, 1082  
 partial 1081  
 Random close packing 617, 1082  
 Rapid  
 cooling 642, 687, 709–10, 712, 721, 730–1, 783, 800, 977, 1012, 1039, 1043, 1062, 1076–9, 1082–3  
 growth of stepped surface 564, 572–4, 588  
 nucleation 573, 607, 623, 628, 632, 637–8, 689, 696, 810, 821, 1072, 1089  
 solidification 630–1, 633–4, 640, 644, 657, 666, 695, 1059, 1076–8, 1083  
 Ratchet mechanism 938–9, 941–4, 950  
 Rate-determining process 816, 977, 990  
 Reaction  
 bainite 799, 1090, 1092, 1095–6  
 continuous 759, 761, 779, 819, 821–2  
 crystallization 1084  
 discontinuous 767, 772, 797, 813, 815–16  
 disordering 819  
 eutectic 633, 666, 692  
 eutectoidal 797–9  
 isothermal 718, 730–1, 749, 752, 782, 797, 833, 977, 1069  
 martensitic 712, 800, 802, 961  
 massive 712, 716, 802  
 monotectic 633, 643, 684  
 order–disorder 1046  
 ordering 736, 818–19, 821, 1046  
 peritectic 642–4, 666, 684  
 recrystallization 719  
 sigmoidal curve 782, 838  
 Recrystallization 598, 719, 832–4, 838–48, 850–1, 854–5  
 in situ 834  
 kinetics of 838, 842  
 primary 832, 837–8, 849, 851, 853–4  
 reaction 719  
 secondary 832–3, 844, 846, 849, 851, 853–4  
 tertiary 854  
 textures 846, 850, 854  
 Reflection twin 860, 866, 903, 919–20  
 Relaxation 555, 745, 893, 919, 978, 1082–3  
 time 566, 818  
 Remelting 633, 692, 694  
 Restabilization 1107  
 Retrogression 774  
 Reversion 774, 776, 778, 782, 1111  
 Rheocasting 693  
 Rods 615, 653, 676, 684, 686, 727, 838  
 Rotation twin 860, 866  
 Roughening 561, 563–4, 579, 609  
 Rubber-like (or ferro-) elasticity 1102, 1105–7
- Schmid  
 factor 973, 1104  
 law 912, 981  
 Second-order  
 chemical reaction 823  
 phase change 819, 828  
 Secondary  
 dendrite arm 691–2  
 recrystallization 832–3, 844, 846, 849, 851, 853–4  
 Seeding 704  
 Segregates 694, 722, 741, 779  
 Segregation 557, 641, 665, 673–6, 691, 694, 696, 722, 724, 736, 741, 751, 769, 771, 775, 848–9, 1085, 1096  
 gravitational 647  
 inverse 694  
 macro- 694  
 micro- 691–2, 1077  
 normal 693  
 Self-accommodation 987, 1031, 1040, 1046, 1069, 1108–10

- Semi-coherent  
 boundary 740, 845, 1049–50  
 interface 598, 759, 778, 914, 1000, 1051, 1064
- Shape  
 change 659, 755 7, 759, 764 5, 859, 909 11, 928, 961, 963 4, 966, 979, 987, 1004 5, 1023, 1031, 1040, 1047 8, 1064, 1093, 1095 6, 1099 100, 1107 9  
 ellipsoid 996 9, 1001
- Shape memory 974, 987, 1039 40, 1102, 1105 7, 1109 12
- two-way (or reversible) 1102, 1107, 1109, 1112
- Shockley partial 764, 921, 924, 931, 941, 944 6, 966 7, 1055
- Short-range order 563, 818 19, 821, 1083, 1107
- Short-range ordering  
 composition (chemical) 1083  
 topological 1083
- Shuffle 863 7, 870 81, 885 99, 903 8, 916, 925, 927, 936, 949 50, 962, 967, 993, 1015, 1047  
 structure 875, 878, 885  
 vector 871 2, 874, 878
- Sigmoidal curve 797, 824  
 reaction 782, 838  
 transformation 814, 823
- Simple eutectic 639 40
- Single interface transformation 979, 984 7, 1036, 1102, 1109, 1112
- Site saturation 807, 811 13, 815, 839
- Small droplet technique 554, 624, 638, 641 2, 979, 1066
- Soft phonon mode 1064
- Solid nucleus 623, 625, 635, 637 8, 640, 644, 656
- Solid solution 558, 634 6, 641 2, 644 6, 665, 702, 708 9, 718 23, 726, 729, 731, 736, 742 3, 747, 761, 763 4, 767, 774, 782, 802 3, 814, 818, 828, 885, 1012, 1077, 1080 2  
 binary 734, 736  
 disordered 703, 818, 828, 892  
 primary 644, 1043  
 supersaturated 709, 718, 737, 739 40, 742, 749, 759, 764, 1012, 1084
- Solid liquid slurry 693
- Solid-state diffusion 665, 679, 683, 692
- Solidification 560, 623 7, 629 34, 636, 638 45, 648 9, 651, 653, 656 7, 665 6, 670, 679, 684 6, 688 92, 694, 697, 1083  
 directional 645, 649, 661, 666, 683  
 rapid 630 1, 633 4, 640, 644, 657, 666, 695, 1059, 1076 8, 1083  
 unidirectional 645 8, 670, 679, 685, 690
- Solute diffusion 645, 647, 649, 665, 667, 672, 674, 683, 723, 781
- Sorbite 805
- Source  
 pole 937, 939  
 prismatic 943 5
- Spheroid 687, 693, 751, 813 14  
 oblate 934, 984, 1065 8
- Spherulitic 687
- Spinodal  
 coherent 721, 726, 729 30, 733  
 conditional 828  
 decomposition 718 32, 736 7, 741, 772, 819, 828, 831, 1064  
 strain 1064
- Splat cooling 631, 697, 716, 1076
- Spontaneous  
 decomposition 722, 915  
 formation 722, 928, 931, 933, 936, 968, 1068  
 nucleation 931 2, 968, 1065 6, 1068  
 transformation 704, 708, 966, 968, 983, 1062
- Squeeze whiskers 612
- Stability  
 absolute 660  
 marginal 660 3
- Stabilization 988 90, 1066, 1074, 1085, 1107  
 thermal 988, 990
- Stable nucleus 600, 605, 623, 625, 738 9
- Stacking fault 688, 781 3, 834 5, 855 6, 909, 913, 923 4, 933 4, 937 9, 941, 945 9, 953 4, 963, 967, 1014, 1036, 1039 40, 1055 6, 1059, 1067 8, 1112
- Stair-rod cross-slip 945 6
- Stoichiometric composition 644, 886
- Stored energy 832 3, 835, 838, 840, 981
- Strain  
 Bain 993, 1015 16, 1030  
 energy 584 5, 598, 696, 704, 710 11, 724 6, 736, 739 40, 785, 788, 833, 836, 839, 843, 854, 914, 934 5, 961, 966, 980 1, 984, 986, 1052, 1062 3, 1067, 1095, 1098  
 hardening 832  
 long-range 1040  
 spinodal 1064
- Stresses  
 long-range 975, 1048 51, 1055, 1064  
 thermal 653, 655, 836, 1102
- Striation 653 5, 675, 847, 1054
- Structural unit model 918
- Structure  
 decagonal 1085  
 dendritic 639, 644, 646 7, 673, 690  
 eutectic 640, 684 5  
 eutectoidal 813  
 shuffles 875, 878, 885
- Subcritical nucleus 600
- Subgrain growth 834, 836 7, 841

- Supercooling 579, 623–4, 626, 34, 636, 639, 42,  
 644, 5, 647–50, 656–63, 672, 680–3, 685, 687,  
 689–90, 693–4, 703, 4, 709–11, 713, 726, 730, 751,  
 753, 759, 771, 810, 1062–3, 1073, 1083, 1085  
 constitutional 648, 671–2, 674, 8, 680, 1, 686, 692, 3  
 Supercritical nucleus 577, 632, 935  
 Superelastic effect 894, 1102, 5  
 Superelasticity 974, 1103, 6, 1108, 12  
 Superheating 694, 5, 703, 708, 9, 714  
 Superlattice 619, 818, 21, 823, 828, 859, 867, 877,  
 881, 883, 885, 7, 893, 6, 916, 17  
 structures (SLS) 598, 821, 828, 885, 6, 893, 917,  
 1039, 1047, 1085  
 Superledges 967  
 Supersaturated solid solution 709, 718, 737, 739, 40,  
 742, 749, 759, 764, 1012, 1084  
 Supplementary displacement dislocations 923, 927,  
 949  
 Surface  
 diffusion 564, 6, 569, 577, 588, 596, 598, 603, 606,  
 608, 613  
 disordered 563, 650, 1  
 melting 563, 4, 650, 695, 1077  
 reconstruction 561  
 Surface free energy 557, 60, 563, 626, 8, 630, 637,  
 648, 661, 3, 677, 8, 697, 774, 782, 785, 792, 836,  
 851, 854, 961, 1065, 1067, 1085  
 anisotropic 560, 651, 659, 662, 3, 683, 806  
 isotropic 575, 659, 662, 676, 1085  
 Symmetry 576, 7, 588, 9, 611, 616, 18, 749, 751, 775,  
 819, 822, 828, 860, 1, 867, 8, 872, 880, 884, 5, 887,  
 893, 4, 903, 905, 907, 918, 923, 925, 7, 937, 939,  
 947, 973, 4, 981, 2, 993, 996, 998, 9, 1007, 1015,  
 1027, 1034, 1036, 1039, 40, 1082, 1085, 6, 1111, 12  
 operations 861, 867, 884, 923, 999  
 Sympathetic nucleation 987, 1069, 1098  
 System  
 eutectic 639, 40, 684, 5  
 monotectic 695  
  
*T T T* diagram 703, 708, 797, 800, 802, 805, 814,  
 977, 1078, 1090, 1096  
*T<sub>i</sub>*, temperature 703, 709, 713, 14, 980, 982, 4, 986,  
 1062, 3, 1066, 1077  
 Temperature  
 eutectic 639, 42, 685, 688  
 eutectoidal 805, 810, 1096  
 order disorder 818, 19, 828  
 peritectic 643  
 Temperature gradient 559, 595, 645, 650, 661, 670,  
 675, 677, 80, 682, 3, 686, 690, 692, 695, 810,  
 984, 986  
 at the interface 645, 668, 672  
 in the liquid 644, 5, 674, 677  
 Tertiary recrystallization 854  
 Tetrahedral packing 1082  
 Texture 686, 690, 1, 848, 850, 1, 854, 912  
 recrystallization 846, 850, 854  
 Thermal  
 analysis 633, 702  
 arrest 642, 702, 710  
 conductivity 648, 661, 677, 683, 689, 1076  
 cycling method 745, 7, 769, 783  
 decomposition 595, 597  
 dendrite 656, 661  
 diffusion 647, 9, 657, 668  
 energy 556, 650, 703, 710, 712, 837, 931  
 gradient 653, 655, 661, 677, 8  
 nucleation 712, 742, 747, 914, 1068, 1072, 1077  
 stabilization 988, 990  
 stresses 653, 655, 836, 1102  
 vibrations 556, 776  
 Thermally activated growth 703, 705, 710, 712, 716,  
 914, 977, 1046, 7  
 Thermoelastic martensites 909, 972, 1069  
 Thin-plate martensite 1014, 1021, 2, 1024, 1054, 5,  
 1067, 1099  
 Thixocasting 693  
 Time-dependent growth 662  
 Tin  
 grey 704, 7, 905  
 white 704, 6, 905  
 Tip  
 dendritic 646, 7, 657, 660, 1, 663, 682, 3, 690  
 radius 647, 649, 656, 63, 679, 682, 3  
 Titanium 595, 617, 713, 14, 866, 896, 7, 899, 902, 4,  
 913, 947, 1040, 4  
 Transformation  
 cubic hexagonal 993, 1043, 1055  
 diffusional displacive 757, 8, 961  
 entropy of 648  
 eutectoidal 797, 8  
 inverse 705, 984, 1112  
 isothermal 703, 710, 730, 759, 761, 797, 8, 800, 3,  
 814, 820, 977, 990, 1071, 2, 1090, 1  
 isothermal curves 625, 6, 708, 726, 814, 977, 8  
 martensitic 702, 709, 10, 713, 716, 749, 764, 820,  
 859, 885, 893, 911, 914, 931, 936, 947, 961, 5,  
 977, 9, 982, 984, 1000, 1012, 1036, 1039, 1043,  
 1047, 1051, 1062, 1064, 1069, 1074, 1077, 1098  
 massive 709, 10, 712, 714, 15, 1062, 1077, 8  
 order disorder 818, 19, 828  
 rate 623, 4, 707, 8, 744, 761, 770, 783, 785, 814,  
 821, 1096  
 sigmoidal curve 814, 823  
 single interface 979, 984, 7, 1036, 1102, 1109, 1112  
 spontaneous 704, 708, 966, 968, 983, 1062

- Transient 573, 578 80, 605, 623, 630, 669, 671, 809,  
1083-4  
nucleation 601, 605-6, 623, 626, 708, 807, 809,  
815
- Trivedi and Kurz theory 648, 661
- Troostite 805
- Twin  
annealing 845, 854 6, 922, 924  
formation of 851, 855 6, 875, 914, 928 9, 947,  
952  
nucleation 933, 947  
nucleus 880, 912, 914, 932 4  
reflection 860, 866, 903, 919 20  
rotation 860, 866  
type I and type II 860 3, 865 7, 869 70, 872 7,  
879 83, 886 7, 891 2, 896 9, 903 6, 914 15,  
917 18, 924 5, 937, 947, 954, 995, 999, 1009 10,  
1012, 1040, 1057, 1108  
type I compound 860, 956  
type II compound 860
- Twinning  
Bullough's model 901 2  
double 867, 880 1, 898 900, 904  
on shock loading 953
- Twinning dislocation 866, 903, 907 8, 911, 915 22,  
924 34, 936 41, 943, 948 9, 954, 956, 975, 1053,  
1057  
complementary 921 2, 928 9  
double 916  
extrinsic 918  
intrinsic 918  
partial 919 21, 924, 937  
zonal 916, 919 22, 924 5, 937, 948, 954
- Twinning mode 859, 861 2, 865 7, 881, 884 9,  
891 4, 896 9, 904 8, 916, 921 2, 924 5, 933, 947,  
950 1, 956, 1025, 1035, 1042  
in double lattice structures 862 3, 866, 875 6, 880,  
888, 901, 907  
in single lattice structures 866, 871, 880, 887 8,  
890, 892
- Two-way (or reversible) shape memory 1102, 1107,  
1109, 1112
- Type I and type II twins 860 3, 865 7, 869 70,  
872 7, 879 83, 886 7, 891 2, 896 9, 903 6,  
914 15, 917 18, 924 5, 937, 947, 954, 995, 999,  
1009 10, 1012, 1040, 1057, 1108
- Type I compound twin 860, 956  
Type II compound twin 860
- Ultrafine particles 611, 614-16  
Undercooled melt 645, 647, 649  
Undercooling 617, 623, 628, 630, 639, 642, 649,  
657 9, 661-4, 682 3, 714
- Unidirectional solidification 645 8, 670, 679, 685, 690  
Upper bainite 1089 90, 1092, 1095 7, 1100
- Vacancy  
condensation 652, 654 5, 783  
supersaturation 654 5, 719, 762
- Vapour  
condensation 554, 566, 594, 614, 635, 1079  
deposition 553, 594 8, 616, 1078  
phase epitaxy 594
- Venables' theory 941 2, 944 6
- Wechsler Lieberman Read theory 993, 995,  
999 1001, 1012, 1015 16, 1018, 1028, 1032
- Whiskers 554, 595, 598, 611 14, 652, 709, 914  
proper 611  
squeeze 612
- White tin 704 6, 905
- Wilson cloud chamber 559
- $\lambda$  interface 987
- X-ray 592 3, 612, 615, 652, 708 9, 718, 720 1,  
725, 732 5, 749 50, 763, 774 5, 778, 821 2, 833 4,  
839, 855, 861, 908, 968, 1072, 1080, 1082, 1085
- Zeldovich factor 577 8, 603
- Zener's hypothesis 656, 682, 742, 770 1, 810 11
- Zinnmeister atomic description 600
- Zirconium 595, 702, 713, 897, 899, 903, 1040 3, 1077
- Zonal (winning) dislocations 916, 919 22, 924 5, 937,  
948, 954
- Zone  
columnar 689 90, 693 4  
levelling 670 1  
melting 670  
refining 664, 670 1, 675, 690, 712, 838, 847 8,  
852 3