DAMAGE BOOK

Drenched Book

UNIVERSAL
LIBRARY

OU_166865

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 669/P796 Accession No. 37041

Author Chalmers, Bruce

Title Progress in Metal Physics

This book should be returned on or before the date last marked below.

PROGRESS SERIES

METAL PHYSICS

PROGRESS IN METAL PHYSICS

1

Editor

BRUCE CHALMERS, D.Sc., Ph.D.

LONDON

BUTTERWORTHS SCIENTIFIC PUBLICATIONS

1949

BUTTERWORTHS SCIENTIFIC PUBLICATIONS LTD
BELL YARD · TEMPLE BAR · LONDON, W.C.2

BUTTERWORTH & CO (AFRICA) LTD · DURBAN

BUTTERWORTH & CO (AUSTRALIA) LTD · SYDNEY
MELBOURNE · BRISBANE · WELLINGTON · AUCKLAND

BUTTERWORTH & CO (CANADA) LTD · TORONTO

American edition published by

INTERSCIENCE PUBLISHERS INC
215 FOURTH AVENUE · NEW YORK, 3

Set in Monotype Baskerville type
Printed in Great Britain by THE KYNOCH PRESS, Witton, Birmingham

CONTENTS

		PAGE
FOREWORD		*vii*
ACKNOWLEDGEMENTS		*viii*
1	PROGRESS IN THE THEORY OF ALLOYS *G. V. Raynor, D.Phil.*	1
2	THEORY OF DISLOCATIONS *A. H. Cottrell, Ph.D.*	77
3	CRYSTAL BOUNDARIES *R. King and B. Chalmers, D.Sc.*	127
4	AGE HARDENING OF METALS *G. C. Smith, B.A.*	163
5	HARDENING RESPONSE OF STEELS *E. H. Bucknall, M.Sc. and W. Steven, Ph.D.*	235
6	PREFERRED ORIENTATION IN NON-FERROUS METALS *T. Ll. Richards, Ph.D.*	281
7	DIFFUSION OF METALS IN METALS *A. D. le Claire, B.A.*	306
AUTHOR INDEX		380
SUBJECT INDEX		386

FOREWORD

THE study of the physical properties of metals has developed through a number of stages. The first was that in which the mechanical properties were correlated empirically with the heat treatment to which the metal had been subjected and, sometimes, to the chemical composition. At this stage the successful treatment of metals was an art, in the sense that experience rather than understanding led to the most satisfactory results. The next stage, in which the internal structure of the metal was examined, was based originally on the use of the microscope and it was found that many experimental facts could be explained in terms of effects that were of the right size to be seen under magnifications of less than about two thousand. The development of the x-ray diffraction techniques allowed phenomena of a smaller order of magnitude to be examined, and much of the existing information was found to be comprehensible in terms of the geometry of the crystal structure of the various phases that were visible under the microscope.

More recent development can perhaps best be discussed by a division of the field into what may be termed 'statics' and 'dynamics'. Under the former heading is the study of the conditions which govern the structure of a metal or alloy when it is in thermodynamic and mechanical equilibrium. The theories of the phases that are present in equilibrium and of the elastic constants have made remarkable progress in terms of rapidly developing theory of the part played by electrons in the metal.

Under the heading of 'dynamic' effects we may include both the conditions governing the approach to equilibrium in respect of the phases that are present, in which diffusion plays an important part, and the response of a metal to forces which are sufficient to cause non-recoverable or plastic mechanical deformation.

These and associated subjects have advanced so rapidly that it has become difficult for research workers in one part of the field to remain up to date in other branches. It is the purpose of this volume, which is to be the first of an annual series, to present authoritative reviews of the present state of knowledge in specialized aspects of the field that includes both physical metallurgy and metal physics. It is not intended that any one volume should form a complete textbook on these subjects. It is hoped rather that a few subjects of current interest should be discussed rather fully so as to cover, in the course of several years, all the more important aspects in which progress is being made. In order to make the series reasonably self contained it is proposed that the necessary 'historical' background should be included the first time a particular subject is discussed. Subsequent articles on such subjects will generally only cover the more recent progress.

B. CHALMERS

ACKNOWLEDGEMENTS

THE following illustrations appear by permission of the authors and the sources given below:

American Institute of Mining and Metallurgical Engineers

Figures 1, p. 164; 10, p. 202; 21, p. 268; 22, p. 269; 23, p. 270; 1, p. 314; 9, p. 364; 12, p. 366; 14, p. 368
Plates VII and XV

American Society for Metals

Figures 5, p. 187; 2, p. 238; 19, p. 265

Journal of Applied Physics
Figure 8, p. 360

Institute of Metals

Figure 1, p. 4; 2, p. 7; 3 and 4, p. 11; 5, p. 13; 6, p. 16; 9, 10 and 11, p. 24; 12, 13 and 14, p. 25; 15, p. 26; 17, p. 27; 18 and 19, p. 28; 27, p. 39; 28, p. 42; 29, p. 44; 41 and 42, p. 63; 43, p. 64; 44, p. 66; 4, p. 182; 6, p. 191; 7 and 8, p. 192; 9, p. 193; 11, p. 216
Plates IV and XVI to XXXII

Iron and Steel Institute

Figures 4, p. 241; 5, p. 246; 6, p. 247; 7, p. 248; 8, p. 249; 11, p. 251; 12, p. 256; 16, p. 263; 19, p. 265.
Plates XXXIII and XXXIV

Metal Progress
Figure 15, facing p. 259

Mining and Metallurgy
Plate VII

Nature, London
Plate VI

Philosophical Magazine
Figures 33, p. 55; 34, 35 and 36, p. 57; 37, p. 58; 38 and 39, p. 59; 40, p. 60

Royal Society

Figures 7, p. 18; 8, p. 19; 30, p. 49; 31, p. 50; 32, p. 51; 45, p. 68; 46, p. 69; 47 p. 70; 7, p. 82; 2, p. 139
Plates I and II

Physical Review
Figures 4, p. 144; 5, p. 147; 2, p. 329; 17, p. 376

University of Illinois
Plate V

1

PROGRESS IN THE THEORY OF ALLOYS

G. V. Raynor

THE PERIOD between the two world wars was one of intense activity in the development of the theory of alloys, and it was during the first part of this period that the use of x-ray crystal structure methods laid the foundations for the progress which followed. Careful systematic experimental work on the constitution of many alloy systems, including several of only theoretical importance, combined with the results of crystal structure determinations on the intermetallic phases, enabled important generalizations to be made. The correlation of these with the results of purely theoretical work on the physical nature of the metallic state showed that a consistent theory of alloy formation could be developed for the relatively simple monovalent metals copper, silver, and gold. More recently progress has been made in applying the electron theory of metals to alloy formation in metals of valency greater than unity, and to the structure of intermetallic phases in general.

The term 'valency' is generally used in metal physics to denote the number of effectively free electrons per atom which are contributed to the metallic structure. In its chemical reactions copper may show a valency of two or one, that is, two electrons may be lost, leaving an ion with twenty seven electrons and two positive charges, or only one electron may be lost, the resulting ion having twenty eight electrons and one positive charge. In the case of metallic copper, however, there is abundant evidence, from both physical and metallurgical sources, that each copper atom contributes to the structure one relatively free electron, and that the sites of the crystal lattice are occupied by singly charged ions. Cohesion arises from interaction between the ions and the free electrons, while the latter, which can be regarded for many purposes as forming an 'electron gas' within the metal, are responsible for such properties as electrical conductivity. It is becoming increasingly clear, however, that the electron theory forms only a part of alloy theory; many other factors have to be considered which may considerably modify the purely electronic effects. Prominent among these other factors are the relative sizes of the atoms of the component

metals of an alloy, the differences in electrochemical properties between the components, and the general nature of the binding forces in the solid state. Progress continues to be made in assessing the effects of these factors, the interplay of which makes the whole problem complicated and generally not amenable to mathematical treatment. At the present stage of the subject, however, it is possible to make limited predictions with regard to the forms of certain binary and ternary equilibrium diagrams, and the general theory of ternary equilibrium is being clarified.

The general theory of the formation of alloys, and the fundamental atomic and electron theory which forms its essential background, have been comprehensively reviewed in several recent publications, particularly by HUME-ROTHERY, to whom much of the development of this subject has been due.[1-4] Since this material is readily available, the present review will deal mainly with recent developments and with certain special aspects of alloy theory. Owing to the immense scope of the subject, no claim to completeness can be made, and many interesting topics will have to be excluded. It will be assumed that the reader is familiar with the modern physical conception of a metal as a regular array in space of metallic ions, with the valency electrons free to move and thus forming an 'electron gas' throughout the metal as a whole.

PRIMARY SOLID SOLUTIONS

Solutions in Copper, Silver, and Gold—In most cases, when one metal dissolves in another it replaces atoms of the solvent metal on the space lattice of the latter, forming what is known as a substitutional solid solution. One of the most important generalizations made by Hume-Rothery was that in copper and silver alloys the extensive formation of primary solid solutions was unlikely if the sizes of the atoms of the solvent and solute metals differed by more than approximately 15 per cent of that of the solvent.[5] If the difference was within this limit, the size factor of the solute was termed favourable, and if outside, unfavourable, with respect to a given solvent. Thus cadmium is within the zone of favourable size factors with respect to silver but outside the corresponding zone for copper; it dissolves extensively in silver but only to a very limited extent in copper. The generalization is discussed in detail, with many other examples, by HUME-ROTHERY.[1,5]

As a measure of the size of an atom, the interatomic distance in the crystal of the corresponding element was originally chosen,[5] since it was

rightly argued that this would be a more fundamental characteristic of the metal than the atomic radius proposed by GOLDSCHMIDT.[6] Goldschmidt's atomic radii were in many cases deduced from the apparent sizes of atoms in alloys, where, however, they are affected by many incidental factors to be discussed later.

For the face-centred and body-centred cubic metals, each atom is equidistantly surrounded by twelve and eight close neighbours respectively, and there is no difficulty in assigning interatomic distances. For close-packed hexagonal metals, however, there are two sets of six neighbours, lying at slightly different distances, for each atom. In these cases, and for other structures in which more than one interatomic distance has to be considered, the closest distance of approach is taken as a measure of the atomic size. Difficulties arise where the metal is incompletely ionized in the metallic state, since it may be fully ionized in an alloy. Its effective atomic size for alloy formation will then be smaller than the interatomic distance in the crystal of the element. Corrections may be applied in these cases to obtain an approximate value of the fully ionized atomic diameter.[5]

Comparison of the atomic diameters of a given solvent metal and a given solute gives a useful guide in deciding whether extensive solid solutions will be formed. Replacement of an atom of the solvent lattice by an atom of approximately the same size will be easier than replacement by a larger atom. In the first case the local distortion produced will be small, and in the second considerable. Detailed examination of the problem shows, however, that it is insufficient to specify the size of an atom by a single constant. Not only the size of the atom, but also the amount of this space occupied by the underlying ion, must be considered.[7,8] Examples of the importance of ionic sizes, of which PAULING[9] and ZACHARIASEN[10] give suitable lists, will be given later in this section. In many cases it is also necessary to take into account the mutual effects of the component metals on each other.[11]

The formation of a solid solution over a wide range of composition will, however, only take place if certain other conditions are also favourable.[5] Where the alloy components differ considerably in electrochemical character, stable intermediate compounds tend to be formed at the expense of the primary solid solution. This becomes clear when the free energy principles of alloy formation are studied. *Figure 1* represents hypothetical free energy curves at a given temperature for

the solid solutions of lead, tin, germanium, and silicon in magnesium, and for the compounds formed. At equilibrium, the free energy for any composition is a minimum. The compositions of the solid solution and the compound in equilibrium are therefore represented by the points of contact of the tangent common to the two curves. Any alloy with a composition between these points has a lower energy if it is a mixture of two phases than if it were to exist as either homogeneous phase. As the compound becomes more stable and the corresponding free energy curve deeper, the point of contact of the common tangent with the solid solution curve moves progressively to lower solute concentrations. Thus a wide range of solid solutions is formed only in the absence of marked electrochemical differences between the components.

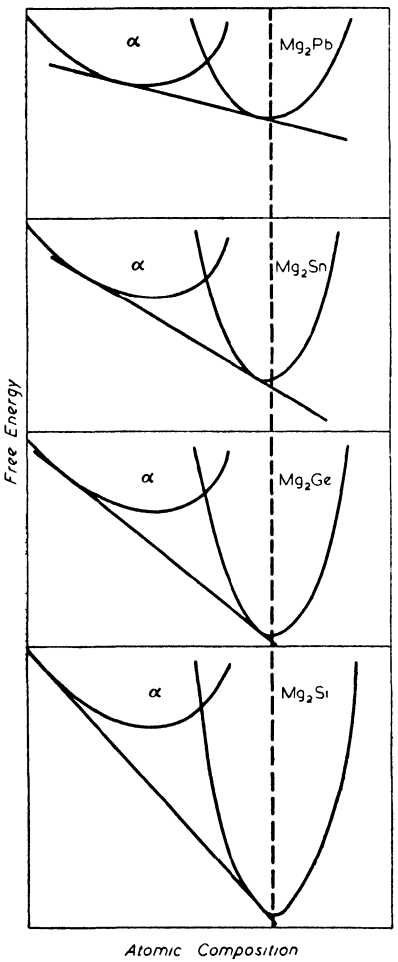


Figure 1. Hypothetical free energy curves at a given temperature for the solid solutions of lead, tin, germanium, and silicon in magnesium and for the compounds formed

It is much more likely that a metal of low valency, such as silver, will dissolve an appreciable amount of a metal of higher valency, such as magnesium, than the reverse (relative valency effect).[5] When magnesium dissolves in silver, atoms with one valency electron each are replaced by atoms with two valency electrons each, and the average number of electrons per atom rises. Conversely, solution of silver in magnesium decreases the electron/atom ratio. Since the electrons are responsible for binding the structure together, it is advantageous to have an excess rather than a deficit of electrons. The binding forces in alloys are more fully discussed later.

For many metals of favourable size factor with respect to silver and copper the solubilities depend intimately on the valency of the solute, being greatest for divalent solutes and decreasing with increasing solute valency. Similarly the liquidus and solidus curves fall more steeply with increasing solute valency. These observations have considerable theoretical significance. From a systematic study of the solid solubilities in copper and silver of the elements of the B sub-groups of the periodic table, Hume-Rothery showed that the maximum solid solubility corresponded to a valency electron/atom ratio of approximately 1·4.[5] In these alloys, the effect of size and electrochemical factors is small. Thus the divalent elements zinc, cadmium, and mercury dissolve in silver up to a limit of approximately 40 atomic per cent, while in the systems copper-aluminium, copper-gallium, silver-aluminium and silver-indium the solubilities of the trivalent solutes are approximately 20 atomic per cent. The solubilities of the tetravalent elements silicon, germanium, and tin are of the order 12 to 13 atomic per cent in copper-silicon, copper-germanium, and silver-tin alloys—a figure which is close to the value 13·3 atomic per cent calculated for a valency electron/atom ratio of 1·4. The solubility of tin in copper, where the size factor is on the borderline of the favourable zone, is, however, only 9·26 atomic per cent. Thus although there is a general tendency for maximum solid solubility in this class of alloy to occur at an electron/atom ratio of 1·4, other factors such as the relative atomic diameters and lattice distortion effects are operative in determining the exact values. It has been shown by HUME-ROTHERY[12] that for solute elements whose size factors are not unfavourable with respect to copper, the mean lattice distortion produced at the solubility limit affects the corresponding electron/atom ratio relatively slightly provided a total expansion of 0·07 to 0·09 Å is not exceeded (copper-zinc, copper-gallium, copper-aluminium). For greater expansions there is a sharp fall in the electron/atom ratio at saturation (copper-tin, copper-indium). For solutes of unfavourable size factor the distortion never reaches this limit, and the intense local effects round solute atoms must be taken into account.

The solid solubilities of many metals in copper, silver, and gold have recently been reviewed by OWEN,[13–16] who has summarized the electron/atom ratios at maximum solubility as in *Table I*, in which the size factors of the solutes are also shown.

Table I. Electron/Atom Ratios at Maximum Solubility

Alloy system	*Electron/atom ratio at maximum solubility*	*Size factor*	*Alloy system*	*Electron/atom ratio at maximum solubility*	*Size factor*
Ag-Cd	1·428	*f*	Ag-Ge	1·29	*b*
Ag-In	1·404	*f*	Ag-Sb	1·29	*f*
Ag-Al	1·403	*f*	Cu-As	1·27	*f*
Ag-Zn	1·40	*f*	Cu-Sn	1·27	*b*
Cu-Al	1·40	*f*	Au-In	1·255	*f*
Cu-Ga	1·39	*f*	Au-Ga	1·25	*b*
Cu-Zn	1·38	*f*	Cu-Sb	1·24	*b*
Ag-Ga	1·36	*b*	Cu-In	1·22	*b*
Ag-As	1·35	*u*	Au-Sn	1·205	*f*
Cu-Ge	1·34	*f*	Au-Hg	1·20	*f*
Ag-Sn	1·335	*f*	Au-Ge	1·09	*b*
Au-Cd	1·33	*f*	Au-Sb	1·045	*f*
Au-Al	1·31	*f*	Cu-Cd	1·02	*u*
Au-Zn	1·30	*f*	Cu-Ag	1·00	*b*

f = favourable *b* = borderline *u* = unfavourable

From silver-cadmium to copper-zinc the size factors are favourable, and the electron/atom ratio changes from 1·43 to 1·38. In the favourable series copper-germanium to gold-zinc the electron/atom ratio changes from 1·34 to 1·30. For the silver-antimony and copper-arsenic systems, in spite of favourable size factors, the ratios are respectively only 1·29 and 1·27, due to the electrochemical effect. For copper and silver, therefore, the maximum solid solubilities for solutes of favourable size factor vary between comparatively wide limits. Alloys with gold as solvent show much lower electron/atom ratios at saturation; the value for gold-cadmium is 1·33, and that for gold-tin 1·205. The general rule is for the electron/atom ratio at saturation to decrease, for the two long periods, as the valency of the solute increases from two to five. The exception to this rule is the series of alloys from copper-cadmium to copper-antimony, to which further reference is made below.

The experimental data may be plotted in the form of *Figure 2*, which brings out the relationship between the maximum solid solubility and the solute valency, and demonstrates the relatively lower saturation values for gold. The difference in maximum solubility for the two series in which the solute is in the same period as the solvent (copper-gallium

to copper-arsenic, silver-indium to silver-antimony) is almost identical. This illustrates that the valency principle is more clearly operative

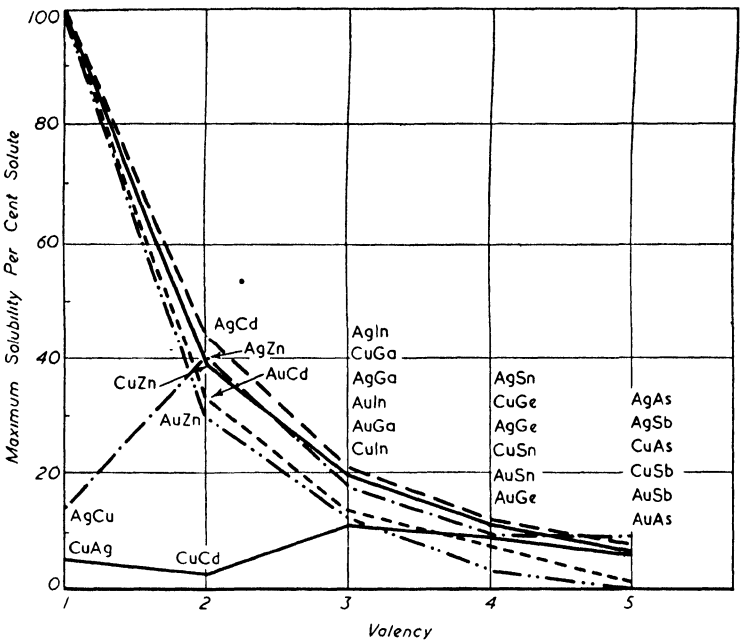


Figure 2. The relationship between maximum solid solubility and valency

in the absence of factors due to the unequal sizes and different electronic structures of the constituent ions.

In special cases a closer insight into the effects of differences in atomic size between the solvent and solute metals may be obtained,[16] and it is found that the maximum solubility is the same for two solute atoms of the same valency if, when dissolved in the same solvent, the ratio of the atomic diameter of one solute to that of the solvent is approximately equal to the ratio of the atomic diameter of the solvent to that of the other solute. Similar results follow if, when the solutes are dissolved in different solvents, the ratio of the diameter of the solute atom to that of the solvent atom is the same in both cases. The following relationships represent the ratios of the atomic diameters; thus cadmium/gold $\simeq$ gold/zinc, and the solubilities of cadmium and zinc in gold are very similar; also gallium and indium dissolve to the same extent in gold, where gold/gallium $\simeq$ indium/gold. For solutions of germanium in copper and tin in silver, in which the maximum

solubilities are again equal, copper/germanium ≏ silver/tin, while similar considerations apply to the silver-indium and copper-aluminium systems, where indium/silver ≏ aluminium/copper. Comparison of the solubilities of arsenic in copper and tin in gold suggests that the same principle holds even for solutes of different valency, provided this lies between four and five.

The solubility limit is thus sensitive to exact relative size relationships between the constituent atoms. It is interesting to note that, for di- or trivalent solutes, solute atoms of greater or smaller diameters than the solvent atom respectively expand or contract the lattice, whereas for tetra- and pentavalent solutes an atom of smaller diameter than that of the solvent may expand the solvent lattice.[13,17] As the solute valency increases, valency factors become more important than size effects in determining the distortion of the solvent lattice. These important effects are discussed later. The maximum solubility, however, appears to be more closely correlated with atomic size differences than with lattice distortion.

Owen explains the lower solubilities in gold in terms of the relative sizes of the atoms and ions of copper, silver, and gold.[16] The amount of space occupied by the ion of a metal *i.e.* nucleus plus non-valency electrons, is less than the amount of space occupied by the atom in the crystal structure.[7] The greater the difference, the greater will be the free space occupied mainly by valency electrons. It will be easier to introduce a large foreign ion into a structure with a large amount of extra-ionic space than into one whose ions are practically touching. The gold ion is larger than that of silver, while the lattice spacing of gold is slightly smaller than that of silver. The gold lattice thus has less extra-ionic space than silver and it will be relatively more difficult to introduce foreign ions into it. Conditions in copper are similar to those in silver. This view attributes low solubilities entirely to the unfavourable configuration of the gold lattice. It is probable, however, that the solute atom also plays a part, in that its ionic charge may deform the electronic atmosphere of the solvent ion.[11] Positive ions in inorganic chemical compounds exert a greater deformation the smaller their size and the greater their charge, so that the outer electrons of a large diffuse negative ion are almost equally affected by the positive charge of a small neighbouring ion as by their own nucleus. In effect, electrons are pulled out of the deformed ion and come partially under the control

of the deforming ion. The same process may occur to a lesser extent in assemblies of unlike positive ions, as in alloys. Thus the large gold ion would be more easily deformable than the smaller silver and copper ions, and when alloyed with a metal having a small, highly charged ion, such as trivalent aluminium, the outer electrons of the gold ion may be envisaged as becoming so much under the influence of the deforming ion as to become effectively free. Support is given to this by the existence of the purple compound[18] Al_2Au, which crystallizes in a structure frequently assumed by phases with an electron/atom ratio of 8/3. The extra electron necessary to give Al_2Au this electron/atom ratio very probably comes from the gold ion, under the influence of the aluminium ions. When fewer deforming ions are present the effect will be less but may well make the contribution of gold to the common stock of electrons greater than one per atom.[19] It will then need less solute metal than in the case of copper and silver, which are less deformable than gold, to raise the number of valency electrons to 1·4 per atom. This serves to emphasize that in alloy problems the possible interactions between ions must not be neglected.

The small solubility of cadmium in copper has been explained as due to an unfavourable size factor. This would suggest that the solubilities in copper of the remaining elements of the same period as cadmium would be equally restricted, since the size factors become no more favourable, but in fact the solubilities in copper of indium, tin, and antimony approach progressively closer to the theoretical value corresponding with 1·4 electrons per atom. This apparently anomalous behaviour illustrates the insufficiency of the simple size factor principle, and indicates that the size of the ionic radii must play a part. On passing along the series of metals from cadmium to antimony, the ions of the solutes decrease in size progressively. It becomes progressively easier, therefore, to insert the ions of cadmium, indium, tin, and antimony into the solvent lattice, and this explains why the solubilities become more nearly normal on passing along the series.[8] The same process should occur with the series gold-cadmium, gold-indium, gold-tin, and gold-antimony if the size of the gold ion is the only factor influencing the relatively low solubilities in gold. Instead, the reverse is the case (*Table I*), giving striking confirmation to the suggestion that the lower solubilities in gold are due to deformation of the gold ion by the solute, because, as the size of the solute ion shrinks and

its charge increases, it becomes a more effective deforming agent and more of the electrons of the gold ion become effectively free. Thus, in agreement with experiment, progressively less of the solute will be required to achieve 1·4 electrons per atom through the series from gold-cadmium to gold-antimony.

Interesting fundamental advances in the theory of solid solubility have therefore been made since the discovery by Hume-Rothery of the constant electron/atom ratio of 1·4 at saturation. These advances have been mainly directed towards sorting out the factors which cause deviations from this value, the theoretical significance of which is discussed later.

Solutions in Magnesium, Zinc and Cadmium[20–22]—The solvents discussed above are effectively monovalent and form alloys in which electrochemical effects are usually small. Magnesium, however, is divalent and strongly electropositive. It therefore forms stable intermetallic compounds with the more electronegative elements of Groups IVB, VB, and VIB of the periodic table. With selenium and tellurium the very stable compounds MgSe and MgTe are formed, while Mg_3As_2, Mg_3Sb_2, and Mg_3Bi_2 are formed with Group VB elements. With Group IVB elements Mg_2Si, Mg_2Ge, Mg_2Sn, and Mg_2Pb are formed. In all these cases the primary solid solution is restricted by compound formation; this restriction is the more marked the more stable the compound (see *Figure 1*). The solubility of monovalent metals in magnesium is restricted by the relative valency effect, except for lithium, where the relatively large solubility can be explained in terms of chemical similarity.[1]

Only Group IIB and IIIB elements, therefore, can be expected to dissolve to any extent in magnesium. Of the solute metals available for examining the influence of size factor effects without interference from compound formation or the relative valency effect, and for which the equilibrium diagrams are accurately known, cadmium, indium, and thallium are of favourable size factor with respect to magnesium. Aluminium is borderline, and zinc and gallium are unfavourable. In agreement with the general size factor principle, cadmium, which also has a close-packed hexagonal lattice like magnesium, forms a continuous series of solid solutions.[23] The characteristic changes in the axial ratios of the alloys may be explained theoretically; superlattices are formed at low temperatures. Indium and thallium have

maximum solubilities of 19·4 and 15·4 atomic per cent respectively, while aluminium dissolves to 11·6 atomic per cent. The borderline nature of the size factor for aluminium is manifested in the marked decrease of solubility with falling temperature. This is characteristic of borderline size factors, so that if it is desired to develop age hardening alloys the solute or solutes should not have too favourable a size factor with respect to the solvent. A further consequence of the borderline size factor is that, though the liquidus curve is practically the same as for the magnesium-indium system, the solidus curve has a much steeper fall. Solidus curves in general appear to be more sensitive to size factor effects than liquidus curves. The solubilities of zinc and gallium are restricted, and both the liquidus and solidus curves fall steeply. For the trivalent solutes, results are summarized in *Figures 3* and *4*, in connection with which it should be pointed out that solutes of the third long period, such as thallium, always give an abnormally large liquidus and solidus depression. A general survey of magnesium alloys confirms that, as in copper and silver alloys, increasing solute valency causes a steeper fall of the liquidus and solidus curves, but the intervention of compound formation at Group IVB precludes a complete survey.

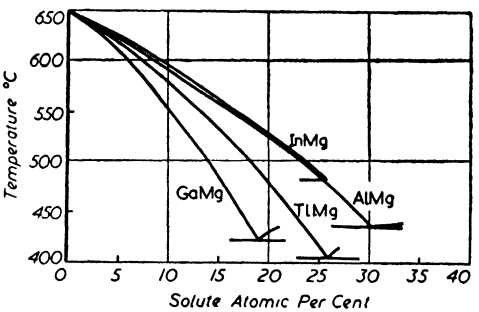


Figure 3. Liquidus curves for the systems Ga-Mg, Tl-Mg, In-Mg, Al-Mg

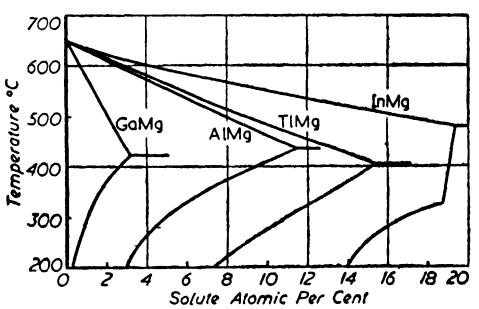


Figure 4. Solidus and solid solubility curves for the systems Ga-Mg, Tl-Mg, In-Mg, Al-Mg

The existence of a relation for magnesium alloys similar to that which limits primary solid solubilities in copper and silver to a composition corresponding to 1·4 electrons per atom is difficult to confirm, since for Group IIB solutes the electron/atom ratio is unchanged, while for Group IVB the compound formation factor interferes. Of the quadrivalent solutes, lead is the element for which compound

formation interferes least. The electron/atom ratio is 2·154 when the maximum solubility of lead in magnesium is reached—a value almost identical with the figure of 2·155 for thallium in magnesium. The values for lead and thallium cannot be compared readily with that for indium, since the abnormally steep solidus depression produced by solutes of the third long period leads to some restriction in solid solubility. Effects for thallium and lead are, however, comparable.

When, therefore, allowance is made for the electropositive and divalent nature of magnesium, it is seen that similar principles hold as for copper and silver alloys. The same principles would be expected to apply to alloys of zinc and cadmium. There is, however, a significant difference between magnesium on the one hand and zinc and cadmium on the other, since the last two metals form only very restricted solid solutions with the elements of Group IIIB. Data are available for the alloys of aluminium, gallium, indium, and thallium in zinc and cadmium, but in no case is the formation of solid solutions appreciable. Even where the size factor is favourable, the highest solid solubility is 3 atomic per cent of aluminium in zinc. This behaviour is due to the difference in the electron energy distribution in the crystals of the elements and is discussed more fully later.

Other Solvents—For solvent metals of higher valency, the interplay of the many factors operative in solid solution formation makes any brief analysis of the situation difficult. It will suffice here to indicate that the size factor principle, the electrochemical factor, and the relative valency effect appear to be of very general application, but that clear valency effects such as are met with in copper and silver alloys are obscured. Beginnings have been made, however, in correlating the solid solutions formed in iron, with special reference to the $\gamma \rightleftharpoons \alpha$ transformation, and interesting results have been obtained. Reference should be made to papers by KORNILOV,[24] HUME-ROTHERY and CHRISTIAN,[25] and ZENER.[26]

Very few elements dissolve to any appreciable extent in aluminium, with the exception of zinc and silver, which both decrease the electron/atom ratio. It is of interest to note that the solid solubility curve for silver in aluminium shows a sharp change of slope at a composition corresponding to 2·6 electrons per atom, which is almost exactly the electron/atom ratio at which the solid solution of zinc in aluminium breaks up into two face-centred cubic phases of different lattice

spacing.[113] Electronic effects may thus be recognized in aluminium alloys but are usually obscured by other factors.

ELECTRON COMPOUNDS

The term 'electron compound' is now well recognized to mean an intermediate phase, in an alloy system, of which the crystal structure and composition are both controlled mainly by the establishment of a certain electron/atom ratio.[1] The recognition of this principle was due mainly to the work of HUME-ROTHERY[27] and to the systematic x-ray work of WESTGREN and his collaborators,[28] and represents a great advance in alloy theory. It has enabled the systematization of many alloy phases, and is of much interest theoretically. Full accounts of the formation of electron compounds at electron/atom ratios of 3/2 (body-centred cubic), 21/13 (γ brass structure) and 7/4 (close-packed hexagonal) may be found in references 1–3.

3/2 Electron Compounds—In many copper, silver, and gold alloy systems, increase of the solute concentration beyond the primary solubility limit

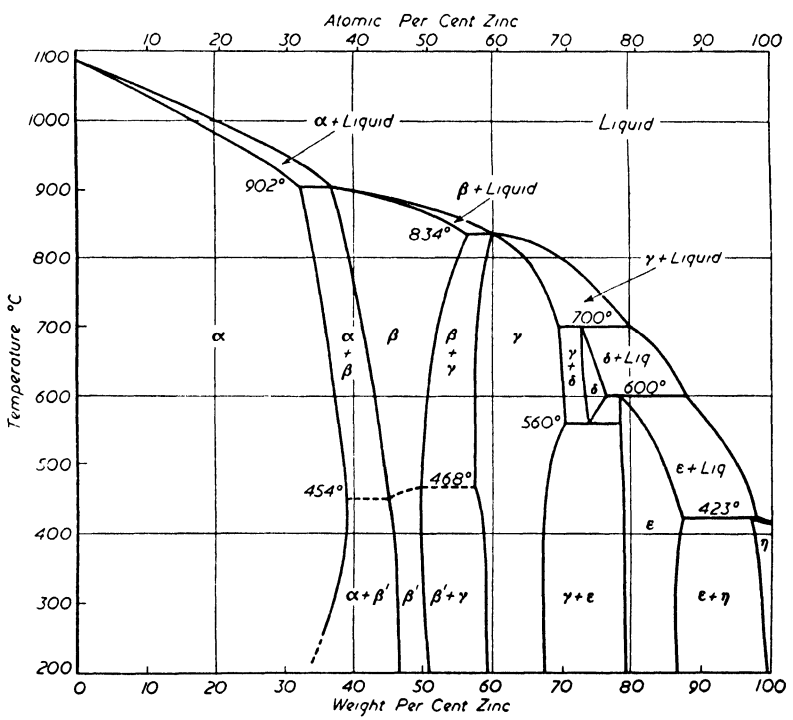
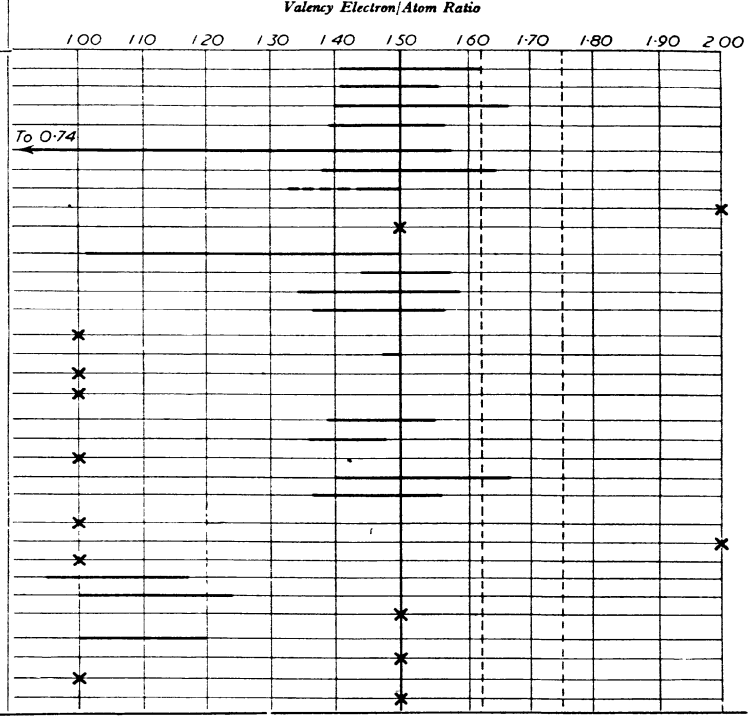


Figure 5. The equilibrium diagram of the system copper-zinc

Table II. Phases with the β brass structure (After Witte)

Description of Phase according to Hansen	*Composition of Phase according to Hume-Rothery*	*Homogeneity Range*		*Homogeneity Range Valency Electrons per Atom*
		Weight %	*Atomic %*	
β Ag-Al	—	6·0–10·0 Al	20·3–30·8 Al	1·41–1·62
β Ag-Cd	AgCd	41·7–56·8 Cd	40·7–55·8 Cd	1·41–1·56
β Ag-Mg	AgMg	13·0–30·0 Mg	39·9–65·5 Mg	1·40–1·66
β Ag-Zn	AgZn	28·0–45·0 Zn	39·1–57·5 Zn	1·39–1·57
γ Al-Co	AlCo	66·2–87·0 Co	47·3–75·4 Co	1·58–0·74
β Al-Cu	$AlCu_3$	83·5–91·0 Cu	68·2–81·1 Cu	1·64–1·38
AlFe	AlFe	—	50·0 Fe	1·50
AlLi	—	—	50·0 Li	2·00
AlNd	—	—	50·0 Nd	1·50
β Al-Ni	AlNi	68·5–81·0 Ni	50·0–66·2 Ni	1·50–1·01
β Au-Cd	AuCd	31·0–44·0 Cd	44·1–58·0 Cd	1·44–1·58
β Au-Mg	—	6·0–15·0 Mg	34·1–58·9 Mg	1·34–1·59
β Au-Zn	AuZn	16·0–30·5 Zn	36·5–57·0 Zn	1·36–1·57
BeCo	—	—	50·0 Co	1·00
γ Be-Cu	BeCu	88·7–87·7 Cu	52·7–50·3 Cu	1·47–1·50
BeNi	—	—	50·0 Ni	1·00
BePd	—	—	50·0 Pd	1·00
β Cu-Ga	Cu_3Ga	20·8–29·2 Ga	19·3–27·3 Ga	1·39–1·55
β Cu-In	—	28·7–36·0 In	18·2–23·8 In	1·36–1·48
Cu-Pd	—	—	38·0–49·8 Pd	1·00
β Cu-Sn	Cu_3Sn	22·5–34·5 Sn	13·5–22·0 Sn	1·40–1·66
β Cu-Zn	CuZn	37·0–56·5 Zn	36·4–55·8 Zn	1·36–1·56
ε Fe-Pt	—	—	35–65 Pt	1·00
HgMg	—	—	50·8 Mg	2·00
Mn_3Si	—	—	25·0 Si	1·00
β Ni-Zn	—	50·0–61·0 Zn	47·3–58·4 Zn	0·95–1·17
CeMg	—	14·8–22 Mg	50·0–62 Mg	1·00–1·24
$CeMg_3$	—	34·2 Mg	75·0 Mg	1·50
LaMg	—	15–21 Mg	50–60 Mg	1·00–1·20
$LaMg_3$	—	34·4 Mg	75·0 Mg	1·50
MgPr	—	85·3 Pr	50·0 Pr	1·00
Mg_3Pr	—	65·9 Pr	25·0 Pr	1·50



Graphical representation of the existence ranges of intermetallic phases with the β brass structure as a function of the valency electron/atom ratio. The characteristic value of the electron/atom ratio for this structure type (1·50) is denoted by a heavy vertical line. The corresponding characteristic values for the two other structure types (γ brass, 1·62; ε brass 1·75) are denoted by broken lines. Phases with unknown homogeneity ranges are distinguished by crosses.

causes a new phase to appear; at still higher solute concentrations the alloys consist entirely of the new phase. This is shown in the equilibrium diagram of the copper-zinc system (*Figure 5*). The second phase frequently has the body-centred cubic structure, either ordered or disordered, but may also have close-packed hexagonal or β manganese structures. The characteristic of these phases is that they are based on an electron/atom ratio of 3/2. Thus the second phase in *Figure 5* is represented as CuZn; those which occur in the alloys of copper with aluminium and tin may be respectively represented as Cu_3Al and Cu_5Sn. *Table II*, which is due to WITTE,[29] summarizes the data on body-centred cubic phases, but also contains details of some phases which are not electron compounds. The systems marked with a cross may be ignored, temporarily, but emphasize that the electron factor is not the only one which leads to this structure. It is seen that the 3/2 electron compounds in most cases have a range of homogeneity on either side of the fundamental composition; this range usually decreases with falling temperature. Some β phases are stable from the melting point to room temperature, but most exist only at high temperatures, and decompose eutectoidally at lower temperatures into the α phase and another phase. The symbol β represents a body-centred cubic 3/2 electron compound. It is important to note that the phases CoAl, NiAl and FeAl can only possess the electron/atom ratio of 3/2 if the transition metals contribute no electrons to the structure. Similarly, transition metal atoms in the electron compounds with electron/atom ratios 21/13 and 7/4 contribute no electrons to the structure. This principle, due chiefly to EKMAN,[30] is well established.

HUME-ROTHERY, REYNOLDS, and RAYNOR[31] have reviewed the 3/2 electron compounds which occur in silver, copper, and gold alloys and have shown that, although the electron/atom ratio is important in controlling their existence, the details of their compositions, homogeneity ranges, and crystal structures are sensitive to other factors, including size factor, temperature, and solute valency. The relevant data are summarized in *Figure 6*, in which the homogeneity ranges are plotted in terms of temperature (vertically) and electron/atom ratio (horizontally). The scales above each drawing refer to atomic per cent of solute, while the numbers prefixed by the plus or minus signs express the percentage size factor derived from the solvent (D_1) and solute

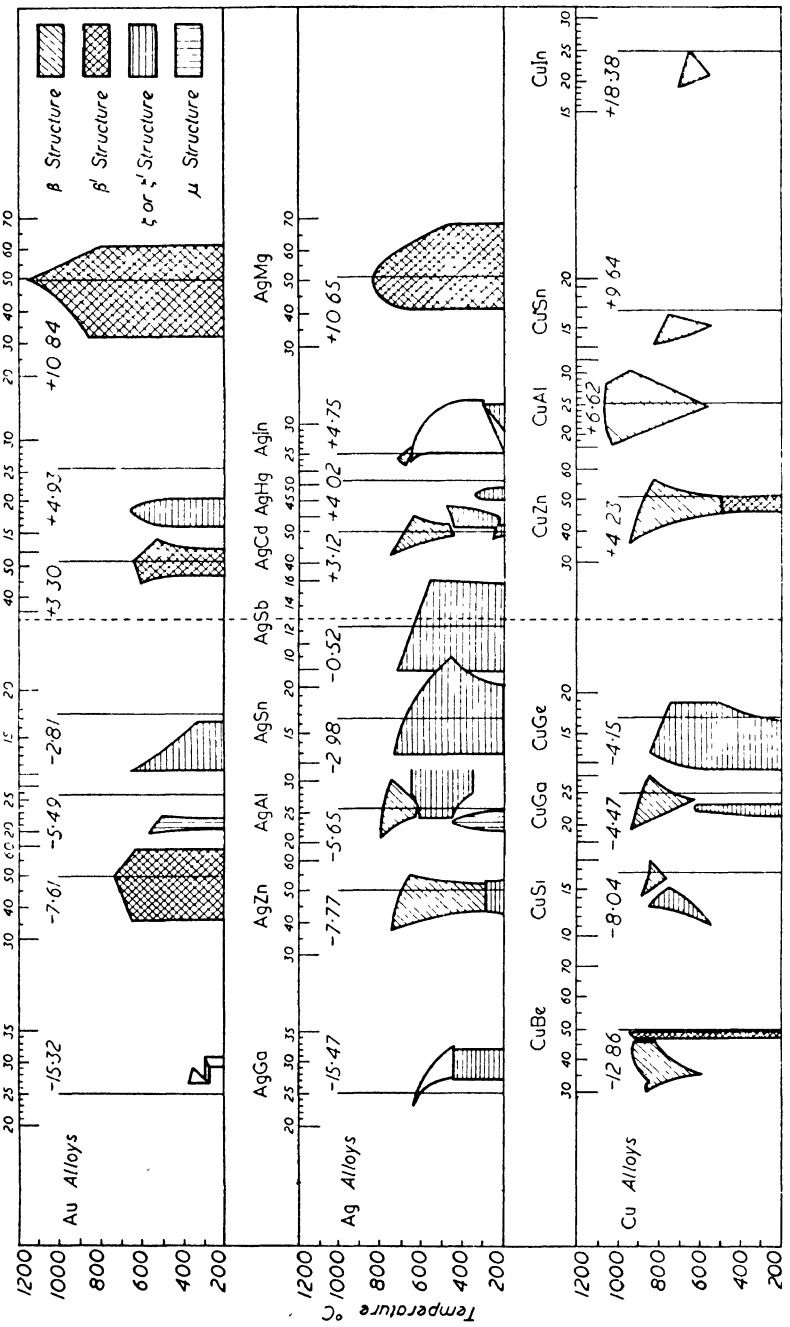


Figure 6. The relationship between homogeneity ranges (plotted in terms of temperature) and electron/atom ratio. The dotted line represents zero size factor

(D_2) atomic diameters, *i.e.* $(D_2 - D_1)100/D_1$. The following general trends are apparent:

1 Increasing solute valency favours the β manganese (μ) or close-packed hexagonal (ʒ) phases at the expense of the body-centred cubic disordered (β) or ordered (β′) phases. The copper-zinc system has both β and β′ phases, the copper-gallium system has a high temperature β phase but a ʒ phase at low temperatures, while in the copper-germanium system only the ʒ phase is stable.

2 Increasing temperature favours the β phases at the expense of the other possible types.

3 Increasing size factor favours the β and β′ phases at the expense of the other possible phases.

4 The general effect of increasing size factor is to displace the β phases to lower electron/atom ratios and to restrict their homogeneity ranges in terms of both electron/atom ratio and temperature.

5 The tendency to form ordered β′ structures follows the sequence gold > silver > copper.

The last observation shows that some other factor in addition to temperature, size factor, and valency is necessary to interpret the forms and stabilities of the 3/2 electron compounds. Detailed study suggests that this factor is the difference in electrochemical character of the components. Gold is more electronegative than copper or silver, which do not differ greatly electrochemically. Thus with electropositive solutes such as magnesium the electrochemical factor is considerably greater for gold than for silver, and somewhat greater for silver than for copper. The tendency to form ordered structures with the Group IIB elements is also in the order Au > Ag > Cu. This tendency cannot be due to size effects, in view of the large difference in stability of the β′ phases for alloys of magnesium with silver and gold, which have almost the same interatomic distances. The tendency to form ordered 3/2 electron compounds is thus greater the greater the electrochemical factor. Many features of these phases may be interpreted according to this principle, and, in general, the higher the electrochemical factor, the more will the electron compound assume the character of true compounds (*i.e.* ordered structures and congruent melting points). This emphasizes a point which is becoming increasingly

clear, that there is no sharp dividing line between the various types of intermediate phases in alloys. In the electron compounds the electron/atom ratio is the most important factor, but the other factors discussed each play their part, and in other compounds may be predominant. The formation of intermediate phases must be regarded as a very general problem of fitting unlike atoms together into a stable structure. The easily classifiable compounds are merely those in which a clearly recognizable factor predominates, and the interplay of all factors must be taken into account.

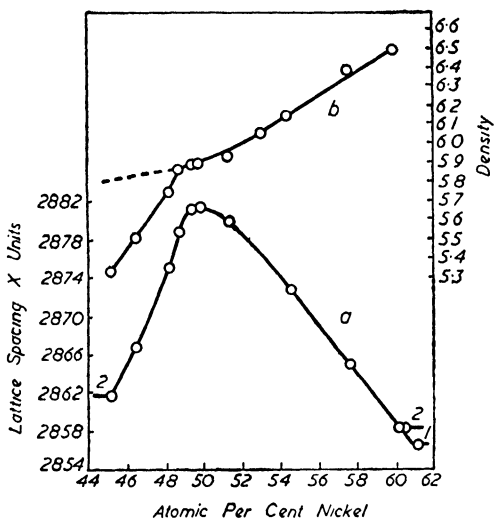


Figure 7. The relationship between a *lattice spacing and* b *density of nickel-aluminium alloys and the nickel concentration. 1 Quenched from 900°C. 2 Two phases*

The importance of the electron/atom ratio is emphasized by the work of BRADLEY and TAYLOR on NiAl[32] and of LIPSON and TAYLOR[33] on some ternary alloys based on this phase. The NiAl alloy containing equal parts of nickel and aluminium has a body-centred cubic structure in which nickel atoms occupy cube centres and aluminium atoms occupy cube corners. With excess of nickel, aluminium atoms are replaced by nickel, and since nickel has smaller atoms and is denser than aluminium, the lattice spacing decreases and the density rises (*Figure 7*). On the other hand, with excess of aluminium, the lattice spacing of NiAl is not increased, nor is the density regularly reduced as would be expected on replacement of nickel atoms by aluminium. Instead, the lattice spacing again falls and there is an anomaly in the density curve. These effects are due to the omission of nickel from some lattice points; the excess of aluminium is obtained by keeping the number of aluminium atoms per unit cell constant, and leaving empty the lattice sites vacated by nickel atoms. The total number of atoms per unit cell falls in such a way as to maintain the number of electrons per unit cell constant. See *Table III.*

It might be argued that spatial considerations would limit the number

Table III. Electron and Atom Ratios for Nickel-Aluminium Alloys

Atomic per cent, Ni	*Number of electrons/atom*	*Number of atoms/unit cell*	*Number of electrons/unit cell*
49·6	1·51	2·00	3·02
48·9	1·53	2·00	3·06
48·4	1·55	1·96	3·04
46·6	1·60	1·89	3·02
45·25	1·64	1·84	3·02
40·0	1·80	1·67	3·00

of large aluminium atoms per unit cell to one, and that the omission of nickel atoms, rather than the replacement of nickel by aluminium, is a consequence of this. This is disproved by the behaviour of the ternary β phase in the copper-aluminium-nickel alloys, which also has a defect structure. The spatial theory requires loss of nickel atoms to begin at 50 atomic per cent of aluminium; the electron theory requires it to begin when the electron/atom ratio 1·5, characteristic of the ideal structure with no defects, is exceeded. *Figure 8*, in which the number of atoms per unit cell is plotted for the ternary alloys, shows that the latter requirement is fulfilled. The NiAl phase therefore prefers the drastic step of omitting atoms from its structure to increasing the number of electrons per unit cell above three. The number of atoms per unit cell is a more fundamental value than the electron/atom ratio, but it will be appreciated that, where there are no defects, the electron/atom ratio is merely the number of electrons per unit cell divided by a constant for a given crystal structure. For a body-centred cube with no defects, the electron/atom ratio 3/2 is strictly equivalent to three electrons per unit cell.

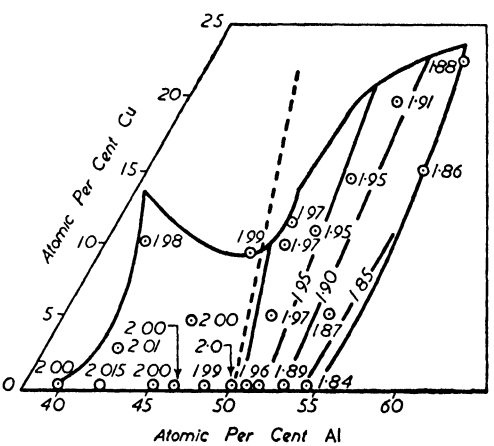
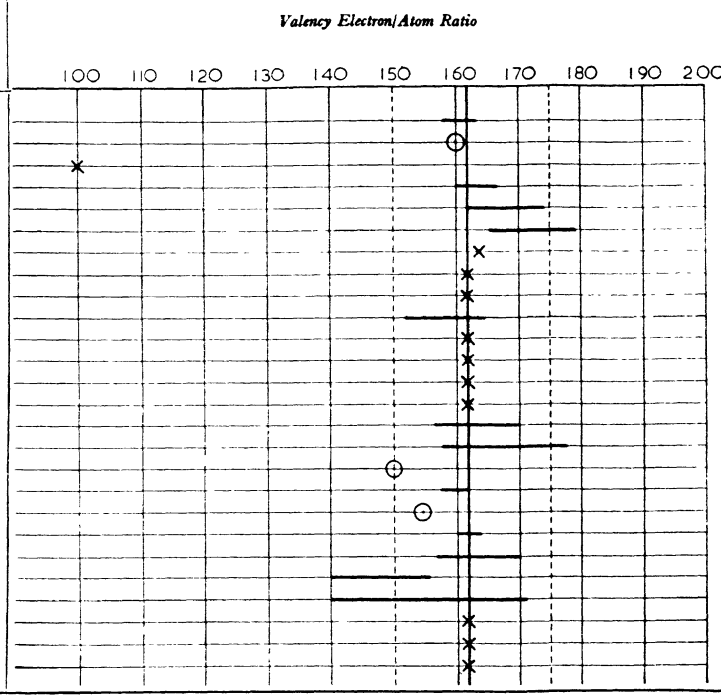


Figure 8. The number of atoms per unit cell for the system copper-aluminium-nickel

21/13 Electron Compounds—In the copper-zinc alloys, the phase Cu_5Zn_8 which occurs at the electron/atom ratio 21/13 has a complex

Table IV. Phases with the γ brass structure (After Witte)

Description of Phase according to Hansen	*Composition of Phase according to Hume-Rothery*	*Homogeneity Range*		*Homogeneity Range Valency Electrons per Atom*
		Weight %	*Atomic %*	
γ Ag-Cd	—	58·5–63·8 Cd	57·5–62·9 Cd	1·58–1·63
γ Ag-Hg	Ag_5Hg_8	74·0 Hg	60·5 Hg	1·60
$AgLi_3$	—	—	75·0 Li	1·00
γ Ag-Zn	Ag_5Zn_8	48·0–55·0 Zn	60·4–66·9 Zn	1·60–1·67
γ′ Al-Cu	Al_4Cu_9	80·0–84·1 Cu	62·9–69·2 Cu	1·74–1·62
γ_1 Au-Zn	Au_5Zn_8	38·0–55·0 Zn	64·9–78·7 Zn	1·65–1·79
$Be_{21}Ni_5$	—	58·0–59·0 Ni	17·5–18·1 Ni	1·65–1·64
$Be_{21}Pt_5$	—	—	19·2 Pt	1·62
$Cd_{21}Co_5$	—	—	19·2 Co	1·62
δ Cd-Cu	Cd_3Cu_8	24·0–34·0 Cu	35·8–47·7 Cu	1·64–1·52
γ Cd-Ni	$Cd_{21}Ni_5$	—	19·2 Ni	1·62
$Cd_{21}Pd_5$	—	—	19·2 Pd	1·62
$Cd_{21}Pt_5$	—	—	19·2 Pt	1·62
$Cd_{21}Rh_5$	—	—	19·2 Rh	1·62
γ Co-Zn	Co_5Zn_{21}	79·5–86·2 Zn	77·8–84·9 Zn	1·56–1·70
δ Cu-Ga	Cu_9Ga_4	31·7–40·0 Ga	29·7–37·8 Ga	1·59–1·76
Cu-Hg	—	—	50·0 Hg	1·50
δ Cu-In	Cu_9In_4	42·3–44·5 In	28·9–30·8 In	1·58–1·62
δ Cu-Si	$Cu_{21}Si_8$	—	18·0 Si	1·54
δ Cu-Sn	$Cu_{21}Sn_8$	32·0–33·0 Sn	20·1–20·9 Sn	1·60–1·63
γ Cu-Zn	Cu_5Zn_8	57·8–70·5 Zn	57·1–69·9 Zn	1·57–1·70
ϵ Fe-Zn	Fe_5Zn_{21}	73·0–80·0 Zn	69·8–77·4 Zn	1·40–1·55
γ Ni-Zn	Ni_5Zn_{21}	72·0–86·6 Zn	69·8–85·3 Zn	1·40–1·71
Pd_5Zn_{21}	Pd_5Zn_{21}	—	80·8 Zn	1·62
Pt_5Zn_{21}	Pt_5Zn_{21}	—	80·8 Zn	1·62
Rh_5Zn_{21}	Rh_5Zn_{21}	—	80·8 Zn	1·62



Graphical representation of the existence ranges of intermetallic phases with the γ brass structure as a function of the valency electron/atom ratio. The characteristic value of the electron/atom ratio for this structure type (1·62) is denoted by a heavy vertical line. The corresponding characteristic values for the two other structure types (β brass, 1·50; ϵ brass, 1·75) are denoted by broken lines. Phases with unknown homogeneity ranges are distinguished by crosses; those with very narrow homogeneity ranges are distinguished by circles.

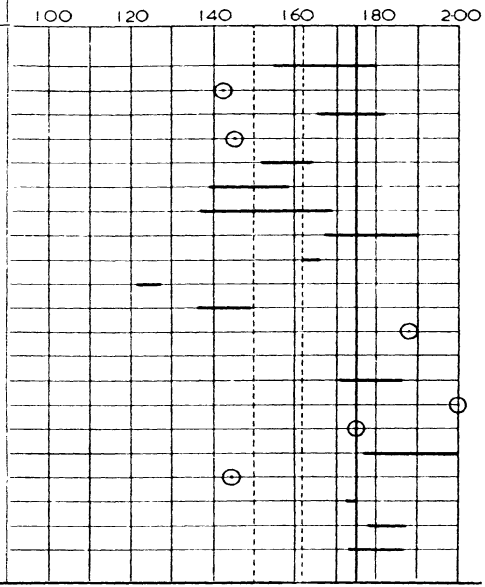
crystal structure with 52 atoms per unit cell. Entirely analogous phases occur in other alloy systems, the data for which are summarized in *Table IV*, which is also due to WITTE.[29] Again transition metal atoms contribute no electrons to the structure, and it should be noted that the range of homogeneity, although usually present, tends to be less than for the 3/2 electron compounds. The existence of the phase $AgLi_3$ (more correctly Ag_3Li_{10}) with the γ brass structure shows that this structure can exist when the valency conditions are not satisfied. On the other hand, the existence of $Na_{31}Pb_8$ shows that the γ brass structure may be formed at the correct electron/atom ratio in systems other than those of copper, silver, gold, or a transition metal. As for 3/2 electron compounds, size factor effects are traceable; although the general structures of Cu_5Zn_8 and Cu_5Cd_8 are entirely analogous, the actual atomic arrangement in the various atomic sites differs,[34] owing to the difference in size of the zinc and cadmium atoms.

The γ brass structure is closely related to other alloy structures. It can be derived from a body-centred cubic structure by stacking together 27 cubes to form a large cubic unit with 54 atoms, and then omitting two atoms to give 52 atoms per unit cell. As described above, the NiAl phase also omits atoms from the structure when aluminium is in excess, and at 45 atomic per cent of nickel the appearance of the alloy resembles γ brass.[32] Since, however, the missing atoms are removed in a random manner, the x-ray pattern is still that of the body-centred cube. It is of great interest to note that, were a γ phase to be formed, its composition would be Ni_6Al_7 (electron/atom ratio 21/13), with a nickel content of 46·1 atomic per cent; this is very close to the composition at which the NiAl phase resembles γ brass, and it is significant that atoms are missing from the structure in both cases. As noted later, the very common nickel arsenide structure also bears a resemblance to the γ brass structure when it has an excess of transitional atoms. Thus for alloy systems of copper, silver, gold, and the transition metals there is usually an attempt to form structures derived from the body-centred cube if this is at all possible.

7/4 Electron Compounds—In the copper-zinc alloys the close-packed hexagonal $CuZn_3$ phase occurs at 7/4 electrons per atom. Analogous phases occur in other systems, as shown in *Table V*. Here, however, there are rather more exceptions to the general rule, since the close-packed hexagonal structure is a common one which could arise from a

Table V. Phases with ε brass structure (After Witte)

Description of Phase according to Hansen	*Composition of Phase according to Hume-Rothery*	*Homogeneity Range*		*Homogeneity Range Valency Electrons per Atom*
		Weight %	*Atomic %*	
γ Ag-Al	Ag_5Al_3	8·5–14·3 Al	27·1–40·0 Al	1·54–1·80
β Ag-As	—	7·5 As	10·5 As	1·42
ε Ag-Cd	$AgCd_3$	65·5–82·0 Cd	64·6–81·4 Cd	1·65–1·81
β Ag-Hg	—	60·0 Hg	44·7 Hg	1·45
δ Ag-In	—	26·8–33·1 In	25·6–31·8 In	1·51–1·63
ε Ag-Sb	—	11·0–16·0 Sb	9·9–14·4 Sb	1·39–1·58
β Ag-Sn	Ag_3Sn	13·5–24·6 Sn	12·4–22·9 Sn	1·37–1·69
ε Ag-Zn	$AgZn_3$	55·3–85·0 Zn	67·1–90·3 Zn	1·67–1·90
γ′ Au-Cd	$AuCd_3$	48·0–51·0 Cd	61·8–64·6 Cd	1·62–1·65
β Au-Hg	—	21·6–27·7 Hg	21·3–27·4 Hg	1·21–1·27
β Au-Sn	Au_3Sn	—	12·0–16·0 Sn	1·36–1·48
ε Au-Zn	$AuZn_3$	72·6 Zn	88·9 Zn	1·89
β Bi-Pb	—	67·0–75·0 Pb	67·2–75·2 Pb	4·33–4·25
β′ Cd-Li	—	—	14·5–29·0 Li	1·86–1·71
β′ Cd-Mg	—	—	75·0 Mg	2·00
η Cu-Ge	Cu_3Ge	—	25·0 Ge	1·75
ε Cu-Sb	—	31·0–39·0 Sb	19·0–25·0 Sb	1·76–2·00
β Cu-Si	Cu_3Si	—	14·5 Si	1·44
ε Cu-Sn	Cu_3Sn	37·7–38·5 Sn	24·5–25·1 Sn	1·73–1·75
ε Cu-Zn	$CuZn_3$	78·5–87·5 Zn	78·0–87·2 Zn	1·78–1·87
η Fe-Zn	—	88·5–94·2 Zn	86·8–93·3 Zn	1·74–1·87



Graphical representation of the existence ranges of intermetallic phases with the ε brass structure as a function of the valency electron/atom ratio. The characteristic value of the electron/atom ratio for this structure type (1·75) is denoted by a heavy vertical line. The corresponding values for the two other structure types (β brass, 1·50; γ brass, 1·62) are denoted by broken lines. Phases with narrow homogeneity ranges are distinguished by circles.

variety of other considerations apart from those of valency. It is of interest to note that the 7/4 electron compound $AgCd_3$ forms unbroken solid solutions, above 300°C, with magnesium.[35] This shows that there is no essential difference in binding between the pure metal and the intermediate phase. Below 300°C superlattices are formed similar to those in the magnesium-cadmium system.

A detailed study of electron compounds shows that the electron/atom ratio is a very important controlling factor in dictating the crystal structure. The theoretical reasons for this are discussed in the next section. At the same time other factors may not only modify the details of the structures taken up but may exert a deciding influence on the ranges of composition and temperature over which the phases are stable. In particular the development of order in electron compounds gives them many of the characteristics of compounds involving molecule formation, and it is impossible to draw sharp dividing lines between the various types. Size relationships may lead to the appearance of structures typical of electron compounds where the establishment of the corresponding electron/atom ratio is impossible or may lead to the formation of new structures, the details of which may be affected by electronic effects. Some examples of this are discussed later.

THE ELECTRONIC THEORY OF ALLOYS

Recent publications[2,3] have dealt in detail with the application of atomic and electronic theory to the study of alloys; only an outline of the subject and some recent developments will be given here.

According to modern theories, moving electrons may be represented as a wave system of wavelength $\lambda = h/mv$, where h is Planck's constant and mv is the momentum of the electron. A free electron therefore has kinetic energy $E = mv^2/2$ or $h^2/2m\lambda^2$, and the energy of an electron can be represented in terms of k, the wave number, which equals $1/\lambda$. The energies of the valency electrons in a metal occupy a series of possible stationary states, which in a crystal of finite size are so close together that they can be regarded as forming a continuum, so that, for a small number of electrons per atom, the graph of E against k is of the form of *Figure 9*. As the number of electrons increases, however, the corresponding k value increases, because only two electrons, with opposite spin quantum numbers, can have energies corresponding to a single energy state; thus N electrons occupy the $N/2$ lowest states. At

a certain wave number k_c for an electron moving in a certain direction in the lattice, the wavelength has a critical value λ_c such that it satisfies the Bragg relation $\lambda_c = 2d \sin \theta$ for diffraction by a certain set of atomic planes upon which it is incident at the angle θ, the interplanar spacing being d. The curve of E against k for this direction of motion then takes the form of *Figure 10*. Electrons with wavelengths slightly

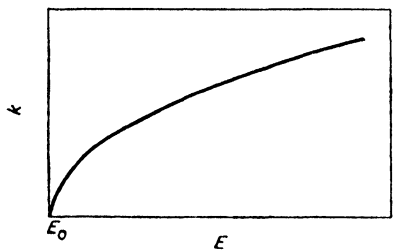


Figure 9. Relation between k *and* E *for free electrons*

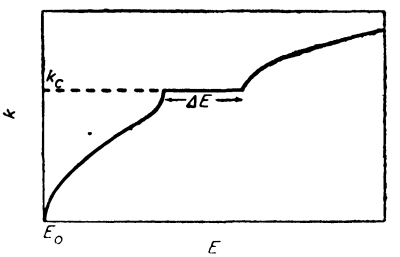


Figure 10. Relation between k *and* E *for an electron moving in a given direction in a periodic lattice*

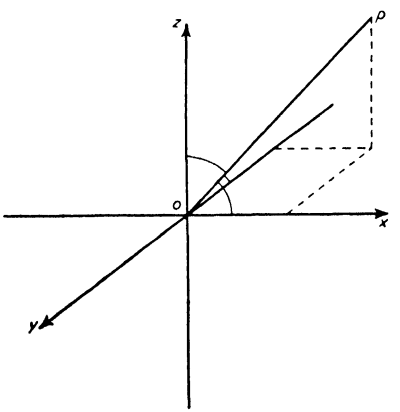


Figure 11. The graphical representation of the critical wave number of an electron moving a lattice

less than λ_c have energies relatively much greater than those with wavelengths slightly greater than λ_c. The energy gap represents a range of energies which an electron cannot possess if it is to be free to move in this direction in the lattice. Similar curves, differing in the position and magnitude of the energy gap, apply for other directions of motion. To describe the energy gaps fully, therefore, some form of three-dimensional plotting is necessary. This is achieved by taking axes k_x, k_y, and k_z mutually at right angles, and plotting the critical wave numbers k_c vectorially. Thus, in *Figure 11*, the length of the line

OP represents k_c, and the direction *OP* represents the direction of motion of the electron with wave number k_0. The critical k_0 values then lie on intersecting planes, in k space, which form a three-dimensional figure with symmetry related to that of the particular metal lattice considered. Across the planes energy varies discontinuously,

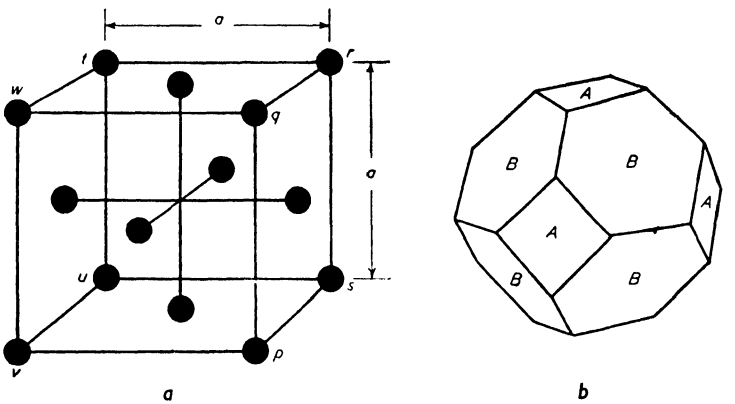


Figure 12. a The face-centred cubic structure and b the corresponding Brillouin zone

but may vary continuously within the enclosed volume. Such enclosed volumes are known as Brillouin zones, and the forms of the first Brillouin zones for the face-centred and body-centred cubic structures are shown in *Figures 12* and *13*. Each can 'contain' a maximum number of

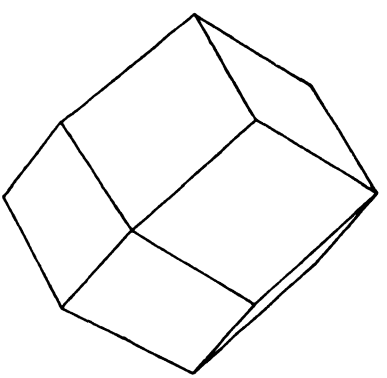


Figure 13. The first Brillouin zone for the body-centred cubic structure

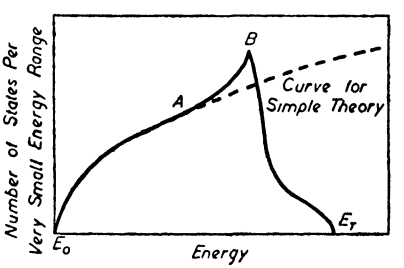


Figure 14. The modification of the distribution of energy states brought about by the presence of a periodic field

two electrons per atom. Energy contours inside such zones are spherical surfaces when the number of electrons per atom is low; as this number increases the energy of the most energetic electron increases and the radius of the spherical surface increases, until, when it begins to approach

the bounding planes of the zone, the surface is distorted from spherical shape. The curve for $\mathcal{N}(E)$ obtained by plotting the number of energy states per unit volume lying between E and $E + \mathrm{d}E$ against E thus takes the form shown in *Figure 14*, where the dotted curve represents the corresponding curve if the zone restrictions are neglected. The area under the zone curve between the limits E_0 and E is the total number of states $\int_{E_0}^{E} \mathcal{N}(E)\,\mathrm{d}E$ with energies in this range. The second

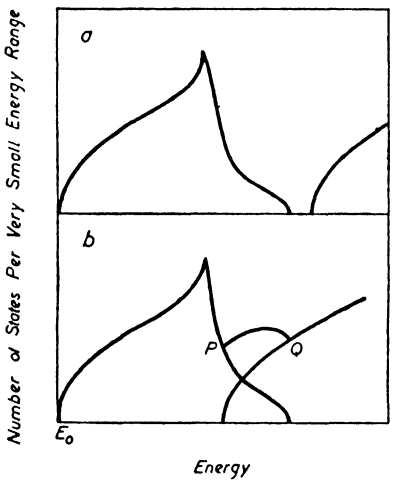


Figure 15. Two possible curves for the distribution of electron energy states for the first and second zones

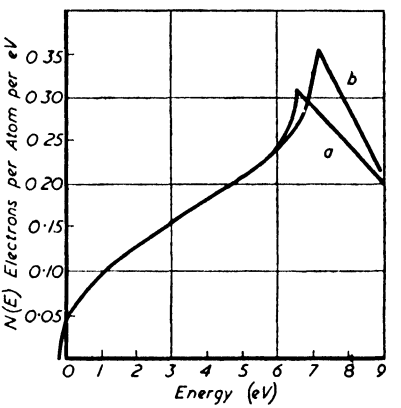


Figure 16. Curves showing number of states per unit energy range as a function of energy a for the face-centred cubic structure, b for the body-centred cubic structure

zone may be separated by energy gaps in all directions from the first zone (*Figure 15a*), but in metals the lowest energy in the second zone is usually lower than the highest energy in the first zone (*Figure 15b*). The zones are then said to 'overlap'.

For the monovalent metals copper, silver, and gold the first Brillouin zone has only half its volume occupied by electrons. Addition of further electrons, *e.g.* by alloy formation, can proceed without hindrance until point B (*Figure 14*) is reached; after this point, owing to the steep fall in the curve, the addition of a small number of electrons causes the energy to rise sharply. It is approximately at this point that the structure would be expected to become unstable with regard to possible alternative structures. In considering equilibrium between two phases, the characteristics of both must be considered. *Figure 16* shows $\mathcal{N}(E)$ curves calculated by JONES[36] for the face-centred and

body-centred cubic structures respectively. When a certain number of electrons per atom is reached, the curve for the face-centred structure *a* falls steeply, while that for the body-centred structure *b* continues to rise. Where curve *a* falls while curve *b* is still rising, the electrons can be accommodated with a lower energy in the zone for the body-centred structure; this therefore represents the point at which the α phase becomes unstable with regard to the β phase in the equilibrium diagram. Calculation gives the electron/atom ratios for α and β in equilibrium as 1·409 and 1·447 respectively, in striking agreement with experiment. The theory only applies to cases of very favourable size factor, since distortion effects are ignored.

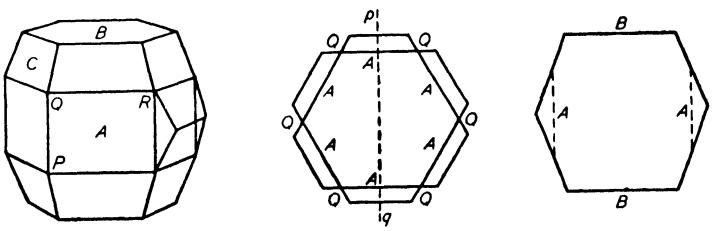


Figure 17. The first Brillouin zone for magnesium a perspective, b equatorial plan, c elevation

According to this theory, alloys tend to crystallize in such a structure that the $\mathcal{N}(E)$ curve is as high as possible. There is thus a tendency for structures to be such that the electrons present are accommodated in the region from E_0 to the point B (*Figure 14*), so that the corresponding Brillouin zone is nearly full. As shown in *Figure 10*, this represents the tendency of the total energy to diminish as much as possible; the energy of electrons just below an energy gap is depressed. This, in brief, accounts for the existence of the 3/2, 21/13, and 7/4 electron compounds. The zones for these structures can contain slightly more than the number of electrons per atom present. This tendency is very general and provides an elegant interpretation of the equilibrium diagrams of copper, silver, and gold alloys and of certain metallic structures, *e.g.* bismuth.

For magnesium, with two electrons per atom and a close-packed hexagonal structure with an axial ratio 1·6225, the theory is more complex.[37] The first Brillouin zone is shown in *Figure* 17 in perspective, equatorial plan, and elevation, and includes the small prism shown on the right of the perspective drawing on each of the vertical A faces.

It can contain two electrons per atom. There are, however, overlaps from this zone, and owing to the asymmetry of the figure, these do not occur symmetrically.[21] For magnesium (and other hexagonal metals with crystallographic axial ratio $c/a < 1{\cdot}63$), overlaps occur in the following order:

a From the incomplete first zone, across A faces, into the small prisms (A overlap).

b From the complete first zone into the second zone at the re-entrant corners Q (Q overlap).

c From the complete first zone into the second zone across B faces (B overlap).

At two electrons per atom, only the A and Q overlaps occur. For zinc and cadmium, with $c/a > 1{\cdot}63$, the corresponding zone is compressed along the hexagonal axis, and B type planes come nearer to the origin than A type planes. Overlaps then occur in the order B, A, Q. The effect of this on the $N(E)$ curves is shown qualitatively in *Figures 18* ($c/a < 1{\cdot}63$) and *19* ($c/a > 1{\cdot}63$). The dotted curves

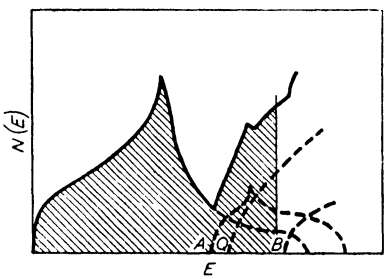


Figure 18. The effect of overlaps at two electrons per atom upon the N(E) *curves for magnesium*

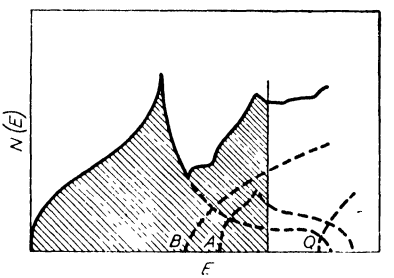


Figure 19. The effect of overlaps at two electrons per atom upon the N(E) *curves for zinc and cadmium*

refer to individual overlaps and the full curve to the resultant energy distribution, while the shaded area represents two electrons per atom. At the limit of the shaded area the curve for magnesium is still rising steeply, while that for zinc and cadmium is almost horizontal. On increasing the electron/atom ratio by solution of elements of higher valency, the energy increase for a given electronic increase is thus greater for zinc and cadmium than for magnesium. This factor explains the different dissolving powers of magnesium, zinc, and cadmium for elements of Group IIIB.

Strictly speaking, the zone theory calculations refer to the absolute zero of temperature. To allow for the effect of temperature, additional

calculations are necessary, introducing entropy terms. JONES has made such calculations for equilibria between the face-centred and body-centred cubic phases in certain copper alloys,[36] assuming that the phase boundaries arise from the difference in the electronic energy of the two phases. Thus the free energy per atom for the α phase may be written:

$$F_a(x_a, T) = f_a(x_a, T) - TS_{ma}(x_a)$$

Here x_a is the atomic concentration, and S_{ma} is the mixing entropy.

$$S_{ma} = -k\{x_a \log x_a + (1 - x_a) \log (1 - x_a)\}$$

$$\text{and} \quad f_a(x_a, T) = U_a(x_a) + \int_0^T c_a \, dT - \int_0^T \frac{c_a \, dT}{T}$$

$U_a(x_a)$ is the internal energy at absolute zero, and c_a is the specific heat at constant volume. Corresponding expressions hold for the β phase. At equilibrium

$$\frac{\partial F_a}{\partial x_a} = \frac{\partial F_\beta}{\partial x_\beta} = \frac{F_a(x_a) - F_\beta(x_\beta)}{(x_a - x_\beta)}$$

and the solution of this equation, subject to certain simplifying assumptions, gives the required phase boundary variation, $U_\beta - U_a$ being known from the zone theory calculations. The results, compared with experiment, are shown in *Figure 20.*

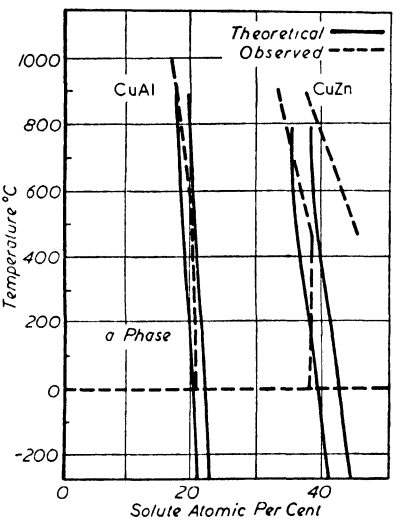


Figure 20. The calculated phase boundaries (full curves) and the observed phase boundaries (broken curves) between the α and β fields of the copper-aluminium and copper-zinc equilibrium diagrams

In the course of the calculation it is assumed that

$$f_a(x_a, T) - f_\beta(x_\beta, T) = U_a(x_a) - U_\beta(x_\beta)$$

The validity of this approximation has been questioned by ZENER,[38] who points out that it is equivalent to assuming that temperature changes alter F_a and F_β in the same manner. Zener also states that Jones' results contain an error of computation.

Zener proposes that account must be taken of the differences in the thermodynamic properties of the two phases and that at temperatures

above the characteristic temperatures of each phase the free energy per atom should be written:

$$F_j(x,T) = U_j(x) + kT\{x \log x + (1-x) \log (1-x)\} - 3kT \log \frac{kT}{h\nu_j}$$

where j refers to α or β, x is considered as a single composition variable, and ν_j is the geometrical mean of all normal modes of vibration. This leads to

$$f_\alpha(x,T) - f_\beta(x,T) = U_\alpha(x) - U_\beta(x) + 3kT \log \frac{\nu_\alpha}{\nu_\beta}$$

so that the difference in the thermodynamical properties of the two phases is determined by ν_α/ν_β. Solution of the equilibrium equations now leads to the following expressions for the phase boundaries, where C_1, C_2, C_3, and C_4 are adjustable constants and x_0 is some arbitrary composition within the range of interest:

$$x_\beta - x_\alpha = \frac{C_1}{C} \text{ where } C = C_2 + \frac{kT}{x_0(1-x_0)}$$

$$\frac{x_\beta + x_\alpha}{2} = x_0 - \frac{C_3 + 3kT \log C_4}{C_1}$$

Exact agreement with the phase diagram of *Figure 5* above the $\beta \rightleftharpoons \beta'$ transformation is obtained by a suitable choice of the four constants. By extrapolating the $(\alpha + \beta)/\beta$ and $\beta/(\beta + \gamma)$ boundaries downwards, an intersection is obtained at approximately 100°C. Thus β brass would become unstable at this temperature if the order/disorder transformation did not intervene.

Further characteristics of the body-centred cubic structure are considered by Zener, with particular reference to the elastic anisotropy, which is illustrated by the fact that the ratio of the elastic moduli measured in the [111] and [100] directions in the β brass crystal at room temperature, $E_{[111]}/E_{[100]} = 8{\cdot}9$.[39] Further, the shear constants C_{44} (referring to shear across the (100) planes) and $(C_{11} - C_{12})/2$ (referring to shear across the (110) planes in the $[1\bar{1}0]$ direction), are in the ratio 18 to 19 according to temperature.[40] Some resistance to shear is provided in all structures by the electrostatic interaction between valency electrons and ions. In the body-centred cubic structure, however, this may be overcompensated by the repulsion arising from exchange interaction between ions with closed shells (*i.e.* no vacant energy states) whose ionic atmospheres are overlapping. This, combined

with the fact that a body-centred cubic packing of balls would have no resistance to a simple (110) $[1\bar{1}0]$ shear, accounts for the low value of $(C_{11} - C_{12})/2$ and the mechanical instability of the lattice with respect to such shears. Analogous behaviour would be expected for all body-centred cubic structures composed of ions with closed shells, and this suggests why this structure is limited to certain places in the periodic table. For alkali metals the ions are too small to overlap appreciably, and repulsion between them at normal temperatures is small. The body-centred cube is thus stable. The non-alkali body-centred cubic metals (except for thallium and barium) all have ions in which the outermost *d* shells are incomplete.

Similar considerations account for the anomalous temperature variations (at relatively low temperatures) of the elastic modulus of β brass. The $(C_{11} - C_{12})/2$ shear constant arises from:

i A positive term due to attraction between valency electrons and ions,

ii A negative term due to repulsion between ions.

The term *i* varies slowly with change in lattice constant, while *ii* varies rapidly. Thermal expansion decreases term *ii* more rapidly than *i*, so that with rising temperature the shear constant increases, instead of suffering the more usual decrease. For β brass itself, the temperature coefficient of $(C_{11} - C_{12})/2$ changes sign at about 200°C owing to the decrease in the coefficient associated with disordering.

Zener's theory suggests that the mechanical instability of the body-centred cubic lattice is more serious the lower the temperature, and it is of very great interest to note that BARRETT has produced a transformation of body-centred cubic lithium into a face-centred form by plastic deformation at − 196°C, in spite of previous statements that the body-centred structure was stable from liquid air temperatures upwards.[41] The transformation is accompanied by audible clicks suggesting abrupt shear movements in small isolated regions round which constraints result; this accounts for Barrett's inability to transform more than about half the sample by the type of straining used. At − 196°C the new structure has a lattice spacing of $a = 4{\cdot}41$Å, as opposed to 3·50 Å for the body-centred structure at the same temperature. The explanation, in terms of Zener's theory, is that since the shear constant $(C_{11} - C_{12})/2$ is small, the amplitudes of thermal vibration in the (110) plane and $[1\bar{1}0]$ direction are large; consequently

the entropy, which is proportional to the logarithm of the amplitude of thermal vibration, is also large. The free energy of the structure $F = U - TS$, so that $\partial F/\partial T = -S$; hence the temperature coefficient of the free energy is high, and the free energy increases rapidly as the temperature falls. At some temperature the free energy for the body-centred cubic structure may be greater than that for the face-centred cubic structure, for which $\partial F/\partial T$ is not so large. A transformation is thus possible, just as certain compositions of β brass change into the face-centred structure on cooling. Such a transformation should be aided by shear in the direction in which the shear constant is low, since not only does it help the atoms to surmount the potential barrier in this direction but it produces an atomic arrangement which is nearly face-centred cubic.

Considerable progress is therefore being made with the mathematical theory of metals and alloys and, by judicious combinations of electron theory and classical thermodynamics, some insight has been gained into the reasons for the assumption of one crystal structure rather than another in specific cases.

OTHER INTERMEDIATE PHASES IN ALLOYS

Electron compounds with electron/atom ratios of 3/2, 21/13, and 7/4 rarely occur outside the comparatively restricted range of alloys with copper, silver, gold, or the transition metals. Considerable attention has been paid to intermediate phases other than electron compounds, and progress has been made in understanding their characteristics.

Several structures which are dependent on size relationships have been discussed by HUME-ROTHERY.[1] These include the interstitial hydrides, nitrides, carbides, and borides of the metals, in which the small non-metallic atom is accommodated in a suitably sized hole in a simple skeleton, usually a close-packed metal lattice, which may itself be somewhat distorted.[42] A large class of substitutional intermetallic compounds crystallizes in the NaTl structure, in which each atom is surrounded by four of its own kind and four of the other component; the arrangement is body-centred cubic and typical of cases where the sizes of the component atoms are closely similar.[43] A face-centred cubic structure, bearing great resemblance to the Cu_3Au superlattice, is frequently assumed where the component atoms are neither too different nor too similar in size. The compounds $NaPb_3$,

$CaPb_3$, $SrPb_3$, $CePb_3$, $CaTl_3$, $CaSn_3$, and $CeSn_3$ have this structure.[44] Certain other structures will now be discussed in more detail.

The $CuAl_2$ Structure—This structure (*Figure 21b*) is assumed by more phases than was formerly recognized.[45] The following compounds crystallize in this structure:

BFe_2	BCo_2	BNi_2	$FeSn_2$	$CoSn_2$	$FeGe_2$
$RhPb_2$	$PdPb_2$	$AuPb_2$	$AuNa_2$	$CuAl_2$	$MnSn_2$

Classification of these compounds in terms of the valencies of the components, or their positions in the periodic table, is impossible. Thus

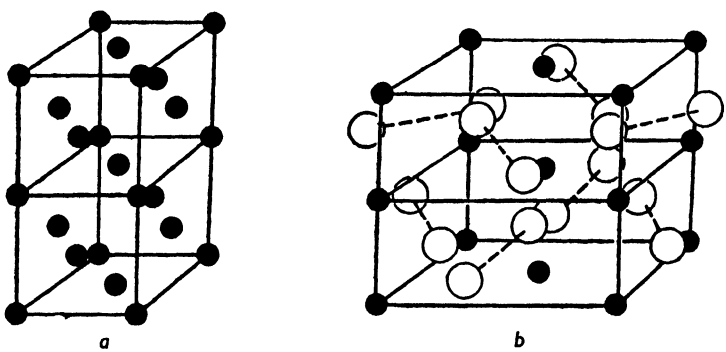


Figure 21. The $CuAl_2$ structure

iron and cobalt can form either the A or the B metal in the formula BA_2; in all cases, however, the A component has a larger atomic diameter than the B metal, and a survey of the structures suggests that for the $CuAl_2$ structure to be assumed, A/B must be of the order of 1·2. The structure may be regarded as derived from two face-centred cubic copper unit cells stacked vertically (*Figure 21a*). If the atoms at the centres of vertical faces are each replaced by a pair of aluminium atoms and the structure elongated horizontally, the $CuAl_2$ structure results. The arrangement is a skeleton of A atoms, in which the smaller B atoms fit. The A atoms form pairs in the [110] direction, and the distance between the atoms of such a pair is shorter than in the metal crystal. This may be taken as evidence of strong bonding, which reinforces the size factor effect. It is to be noted that each atom is surrounded by more than twelve atoms, so that the coordination number is high. This confirms that the size relationships are important. See later.

The Laves Phases AB_2—Many phases crystallize in one of three fundamental types of structure, characterized by the compounds $MgCu_2$, $MgZn_2$, and $MgNi_2$. These are collected in *Table VI.* Interpretation of the whole group in terms of electron rules is impossible, and though there is a tendency for transition metals to be involved this is not universal, and a transition metal may form either the A or the B partner. The partners may come from any part of the periodic table, and the same metal may play different roles in two compounds (*e.g.* KBi_2 and $BiAu_2$). The chief reason for the existence of these phases is that, geometrically, space may be conveniently filled by these three structures when $D_A/D_B = (1{\cdot}5)^{1/2}$ or 1·225, where D represents the atomic diameter. The observed ratios vary between about 1·2 and 1·3. The stability is a reflection of the very high coordination numbers. A high coordination number may roughly be taken to mean that more bonds per atom can be formed than in a metal. In these structures each A atom has twelve B and four A neighbours, giving a coordination number of 16. The coordination number for B atoms is 12, so that in AB_2 the average coordination number is 13·33, whereas the highest a metal can reach is 12. Such a high coordination indicates purely metallic character, and precludes ionic or covalent character. The narrow range of homogeneity of these compounds is due not to any special affinity, but to geometrical factors. This is illustrated by the existence of KNa_2 ($D_K/D_{Na} = 1{\cdot}227$).

Table VI. Compounds Crystallizing in the Three Fundamental Structure Types

$MgCu_2$ *type* (*cubic*)	$MgZn_2$ *type* (*hexagonal*)	$MgNi_2$ *type* (*hexagonal*)
$AgBe_2$ $CaAl_2$ $NbCo_2$ $CeMg_2$	$CaMg_2$ $TiFe_2$ $ZrOs_2$	$CrBe_2$ WBe_2
$TiBe_2$ $LaAl_2$ $TaCo_2$ ZrW_2	$TiMn_2$ $NbFe_2$ $ZrCr_2$	$MnBe_2$ $TiCo_2$
$NaAu_2$ $CeAl_2$ $CeNi_2$ $PbAu_2$	$ZrMn_2$ $TaFe_2$ KNa_2	VBe_2 $ZrFe_2$ } with approx. 7·4% excess of 'B' metal
$TiCo_2$ $ZrFe_2$ $CeCo_2$ $BiAu_2$	$NbMn_2$ $ZrRe_2$ $SrMg_2$	$ReBe_2$ $NbCo_2$ } with approx. 7·4% excess of 'B' metal
$ZrCo_2$ $LaMg_2$ KBi_2	$TaMn_2$ ZrV_2 $BaMg_2$	$MoBe_2$ $TaCo_2$ } with approx. 7·4% excess of 'B' metal
	$CaLi_2$	FeB_2

The three structures are closely related,[46] though as usually illustrated this is not easily seen. They can be regarded as interpenetrating lattices of A and B atoms, and the resemblances are best revealed by considering these separately. The zinc, copper, or nickel atoms are

arranged in a pattern of tetrahedra. The arrangement in $MgZn_2$ is shown in *Figure 22a*. The tetrahedra, in rows and joined alternately at base and point, enclose a hole in which the magnesium atom lies. In $MgCu_2$ (*Figure 22b*) the tetrahedra touch at points throughout,

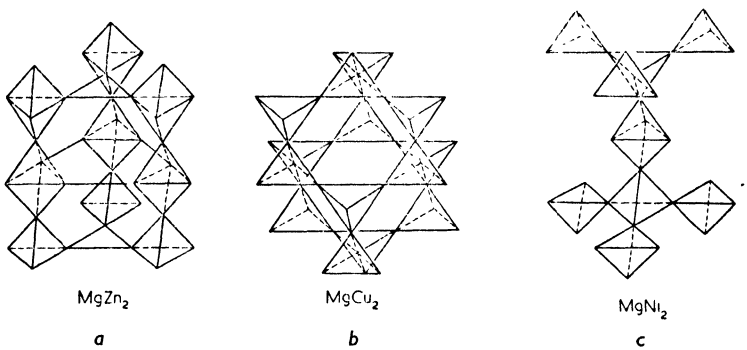


Figure 22. The structure of a $MgZn_2$, *b* $MgCu_2$, *c* $MgNi_2$

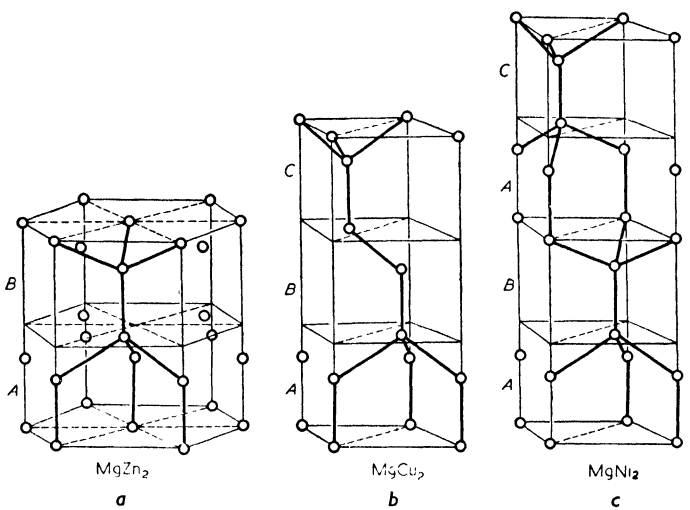


Figure 23. The disposition of magnesium atoms in the structure of a $MgZn_2$, *b* $MgCu_2$, *c* $MgNi_2$

while the $MgNi_2$ arrangement (*Figure 22c*) is a mixture of the other two types. The disposition of the large Mg atoms in the three structures is shown in *Figure 23*. In $MgZn_2$ the arrangement is that of wurtzite (*Figure 29b*); in $MgCu_2$ the arrangement is that of zinc blende (*Figure 29a*). $MgNi_2$ has a mixture of both types. This explains why the axial ratios of $MgZn_2$, $MgCu_2$, and $MgNi_2$ are in the ratio 2 : 3 : 4. $MgZn_2$

can be represented as double layers of hexagonally arranged magnesium atoms, so arranged that one double layer *B* is shifted laterally relative to the other (*Figure 24a*). For $MgCu_2$ the same two double layers occur, but instead of the sequence *ABABAB* a different type of double layer is introduced (*Figure 24b*) and the sequence is *ABCABC*. In the $MgNi_2$ structure the same types of double layers are involved, but the sequence is *ABACABAC*. All three structures are thus fundamentally similar.

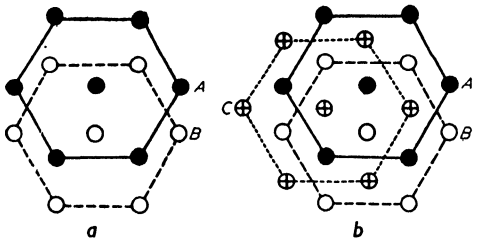


Figure 24. Plan of arrangement of magnesium atoms in the structure of a $MgZn_2$, *b* $MgCu_2$

Though size relationships are the governing factor, electronic factors are also operative. In the system magnesium-nickel-zinc, the ternary phase MgNiZn is isomorphous with $MgCu_2$, in spite of the fact that $MgNi_2$ and $MgZn_2$ occur in the system;[47] if the effective valency of nickel is taken as zero, the electron/atom ratio for $MgCu_2$ and MgNiZn is 4/3. The quaternary compound $Mg_3Zn_2Cu_2Ni_2$, isomorphous with $MgCu_2$, has the same electron/atom ratio also if nickel is effectively zerovalent. Again, the ternary phase MgCuAl (electron/atom ratio 2) is isomorphous with $MgZn_2$ with the same electron/atom ratio.[47]

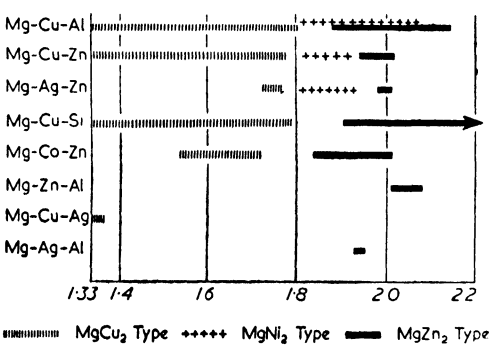


Figure 25. Schematic diagram showing the dependence of the range of existence of the $MgCu_2$, $MgNi_2$, *and* $MgZn_2$ *structures upon the electron/atom ratio for several ternary magnesium alloys*

Further evidence is obtained by the experiments of LAVES and WITTE[48] on the effect of changing the electron/atom ratios of $MgCu_2$ and $MgZn_2$ by alloying (*Figure 25*). The experimental results may be summarized as follows:

1 When copper is replaced by zinc in $MgCu_2$, the $MgCu_2$ structure persists up to approximately 1·8 electrons per

atom, and is then succeeded successively by the $MgNi_2$ and $MgZn_2$ structures.

2 When copper is replaced by aluminium in $MgCu_2$, the $MgCu_2$ structure again persists up to about 1·8 electrons per atom, and is succeeded by the $MgNi_2$ structure up to 2·07 electrons per atom at high temperatures, or 1·88 electrons per atom at low temperatures. For the remainder of the solid solution field up to 2·15 electrons per atom the $MgZn_2$ structure occurs.[49]

3 When zinc in $MgZn_2$ is replaced by silver, the reverse sequence occurs.[50]

4 When zinc in $MgZn_2$ is replaced by aluminium, the $MgZn_2$ structure persists up to 2·08 electrons per atom, above which value phases of type AB_2 do not occur.

The $MgCu_2$ structure thus occurs at a low electron/atom ratio, and the $MgZn_2$ structure when the ratio increases to approximately two. The presence of the $MgNi_2$ structure at intermediate electron/atom ratios suggests that nickel has a valency between one and two, which is improbable in view of the isomorphy of $MgCu_2$ and MgNiZn. It is more probable that $MgNi_2$ structures in these cases are structural intermediates between the other two structures. Thus the transition $MgZn_2 \rightarrow MgCu_2$ involves a periodic 'fault' at every third double layer of magnesium atoms, but the transition $MgZn_2 \rightarrow MgNi_2$ involves less severe faulting of the same type. The process thus occurs in stages and the observation[50] that when the zinc in $MgZn_2$ is replaced by silver intermediate types of structure occur between the main types supports this view. The formation of the $MgZn_2$ structure on cooling certain alloys with the $MgNi_2$ structure in the magnesium-copper-aluminium system[49] is also consistent with it.

Further confirmation of the valency relations is given by the appearance of the $MgZn_2$ structure on adding silicon to $MgCu_2$.[51] It should also be noted that, if the valency of nickel in these compounds were really between one and two, replacement of nickel in $MgNi_2$ by a polyvalent metal should lead to the appearance of the $MgZn_2$ structure; instead the $MgCu_2$ structure appears,[48] as would be expected if nickel were zerovalent. Similarly, addition of silicon to MgNiZn gives the $MgZn_2$ structure, while addition of cobalt to $MgZn_2$ gives the $MgCu_2$ structure, as would be expected if cobalt contributed no valency

electrons. The occurrence of the $MgZn_2$ structure at an electron/atom ratio of 1·9 in the magnesium-silver-aluminium system should be noted.

Figure 26 summarizes the homogeneity ranges of the various structures in the series $MgCu_2$-$MgZn_2$, $MgCu_2$-($MgAl_2$), $MgCu_2$-($MgAg_2$), $MgZn_2$-($MgAg_2$), and $MgZn_2$-($MgAl_2$). The formulae in parentheses denote compositions and do not correspond to actual compounds. Since the

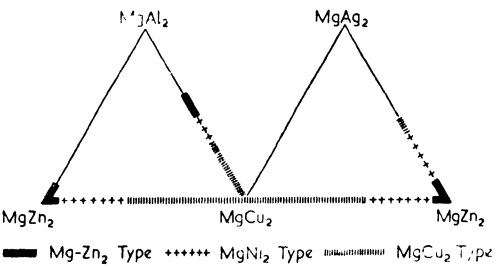


Figure 26. The relationship between the range of existence of the $MgCu_2$, $MgNi_2$, *and* $MgZn_2$ *structure and concentration (molecular per cent) for several ternary magnesium alloys. The atomic radii of magnesium, silver, aluminium, zinc and copper are 1·60, 1·44, 1·43, 1·37 and 1·28 Å respectively, and the left and right side of the figure with regard to the ratio of atomic radii are almost mirror images*

atomic radii of silver and aluminium are very similar, the two triangles are mirror images as far as size effects are concerned. The structural changes in one triangle are not, however, mirrored in the other, confirming that the transformations depend on valency and not size effects.

It is possible for apparently more complex compounds to crystallize in one of these structures, with some of the A atoms replaced by B atoms; thus $FeBe_5$ ($MgCu_2$ structure) may be regarded as $(Fe_{1/2}Be_{1/2})_2Be_4$,[52] while $Cu_4Mg_3Be_2$ ($MgCu_2$ structure) is of the form $MgB_xB'_{2-x}$.[53]

Closely allied to the Laves phases are ternary phases of the type $Mg_3Zn_3Al_2$ and Mg_4CuAl_6, which are isomorphous and, according to Laves, Lohberg, and Witte,[54] may both be regarded as derived from $Mg_{37}X_{63}$, where X is aluminium, copper, or zinc. It is simpler to regard both structures as based on a ratio of twelve large (magnesium) atoms to twenty or twenty one smaller zinc, aluminium, or copper atoms. Similarly the isomorphy of $Mg_3Cu_7Al_{10}$ with Mg_2Zn_{11} is due

to the establishment of a ratio of six large to thirty three or thirty four smaller atoms. The isomorphy of $MgZn_2$ and AlCuMg is mainly a valency effect, probably reinforced by the size ratio effect; it is important to note that, when AlCuMg dissolves zinc, the electron/atom ratio remains constant, while the ratio of large to small atoms deviates from 1 to 2.[55]

This class of compounds shows how structures based essentially on size factor relationships may be modified by electronic considerations. They confirm that the structures of solids are due to the interplay of many simultaneously operative factors.

The Nickel Arsenide Structure—Recent work has considerably enlarged our knowledge of this structure, which is shown in *Figure 27*, and is best regarded as based on a close-packed hexagonal lattice of metalloid atoms, in the octahedral spaces of which are situated the metal atoms. Each atom has six neighbours. Compounds of the type AB with nickel arsenide structures are given in *Table VII.*

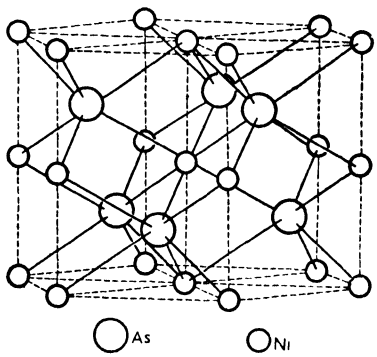


Figure 27. The nickel arsenide structure

Table VII. Compounds of Type AB with Nickel Arsenide Structures

CrS	FeSe	NiTe	FeSn	CoSb	PdSb
CoS	NiSe	CrTe	AuSn	FeSb	PtSb
FeS	CrSe	MnTe	CuSn	NiSb	MnAs
NiS	CoTe	PdTe	PtSn	CrSb	NiAs
CoSe	FeTe	PtTe	NiSn	MnSb	NiBi
		MnBi			

One of the most striking features of this class of alloy, which is formed mainly between electronegative elements and transition metals, and is thus heteropolar, is that there is often a wide homogeneity range which may not include the equiatomic composition. Also several compounds well removed from the composition AB have the same basic structure ($NiTe_2$, Ni_2In, Cu_2In, Fe_2Ge, Co_2Ge, Ni_2Ge, Mn_2Sn).

The γ phase of the nickel-antimony system (46·4 to 54·4 atomic per cent of nickel) has the nickel arsenide structure, and, when nickel is in excess of the composition NiSb, nickel atoms fill up additional holes in the lattice, and may thus be regarded as in interstitial positions.[56] On the other hand, when antimony is in excess, some atomic sites among those occupied by nickel in the ideal structure are left vacant, and the structure becomes of the defect type. The ε phase of the iron-antimony system (42 to 48 atomic per cent of antimony) and the γ phase of the system cobalt-antimony (43·4 to 49·2 atomic per cent of antimony) are both of the interstitial type, with extra metal atoms inserted into holes in the ideal structure.[57] In ternary iron-nickel-antimony alloys, ε FeSb and γ NiSb form an unbroken series of solid solutions;[58] alloys on the FeNi-rich side of a line joining the compositions FeSb and NiSb are of the interstitial type, while those on the Sb-rich side of this line are of the defect type. Solid solutions between ε FeSb and γ CoSb are all of the interstitial type.

The behaviour in the iron-nickel-antimony alloys is strikingly reminiscent of that in the ternary alloys based on NiAl (discussed earlier) and suggests that the modifications in structure with change of composition take place in such a way as to preserve a constant number of electrons per unit cell. If the valency of the transition metal is zero in these three compounds, then the electron/atom ratio is 2·5 for the ideal structure, and since there are two molecules per unit cell, the number of electrons per unit cell is ten. Inclusion of extra transition atoms does not alter this number, while omission of transition atoms, giving a stoichiometric excess of antimony, also leaves it unchanged. This point of view receives support from a determination of the Brillouin zone for this structure by MAKAROV.[59] The zone can contain, for the equiatomic structure, a number of electrons per atom which varies slightly according to axial ratio but is close to 2·5. It should be noted that the nickel arsenide phases without a transitional atom (AuSn, CuSn, Cu_2In) have either the electron/atom ratio 2·5 or ten electrons per unit cell of two molecules, thus supporting the electronic interpretation.

The large number of transition metal compounds which take this structure is probably due to the ease with which the valency of the transition metal can vary. Whereas in NiSb the nickel contributes no electrons, in NiSn it probably contributes one per atom. The

establishment of ten electrons per unit cell for NiS would require nickel to be capable of absorbing electrons. This probably happens, as we shall see later. It is becoming increasingly clear that transition metals may not always exhibit the same valency in alloy systems; the valency shown will depend on the nature of the alloying partner and its relative amount.

We can now understand the existence of Ni_2In. It arises simply by filling up all available interstices in the nickel arsenide structure with nickel atoms. The structure is then complete and contains two indium atoms, two normal nickel atoms, and two interstitial nickel atoms per unit cell.[60] The Cu_2In structure is entirely analogous.

The nickel arsenide structure is closely related to other alloy structures. For nickel arsenide itself, all the octahedral voids in the close-packed hexagonal metalloid lattice are filled by metal atoms. If only half these voids are filled, the cadmium iodide type of lattice results. This accounts for the formation of $CoTe_2$ and $NiTe_2$, which are formed continuously from CoTe and NiTe by partial loss of metal atoms from the structure.[61] If, on the other hand, metal atoms fill additional voids, the Ni_2In type of structure results. We thus have a series of compounds such as $NiTe_2$, NiSb, NiTe, and Ni_2In, fundamentally all based on the same lattice.[62] Such compounds are heteropolar, but the metallic character increases as the proportion of small metal atoms increases.[63,64] At the same time, the axial ratio changes. For simple stoichiometric nickel arsenide phases, $c/a \simeq 1{\cdot}6$. The phase of composition $Ni_{1{\cdot}8}In$ has $c/a = 1{\cdot}233$, while Ni_2In has $c/a = 1{\cdot}228$. A full survey[65] shows that as additional voids are filled by metal atoms the value of c/a approaches more closely to 1·225. This axial ratio gives rise to a pseudo-cubic structure with a striking resemblance to the γ brass structure, a fact which has been used to suggest that γ brass is a heteropolar compound, analogous to Ni^+As'. While it is probable that γ brass phases have heteropolar character, this is likely to be in the opposite sense.[66] In Ni_5Cd_{21}, 42 electrons, supplied by the cadmium, are approximately equally divided between the 26 atoms present, so that each nickel atom has 42/26 more and each cadmium atom 10/26 fewer electrons than normally. Thus the heteropolarity is in the sense $Ni_5'Cd_{21}^+$. If, however, the Brillouin zone interpretation of nickel arsenide phases be accepted, the heteropolarity will be in the same sense as for the γ brass structure.

The tendency to avoid the stoichiometric composition AB depends

on the nature of the metal B.[63] As the number of electrons in the outermost sub-group in the atom of the B metal falls, the structure passes from the defect type (excess of B) to the interstitial type (excess of A). The electrons of the B metal appear to be involved in binding together the skeleton of B atoms. When only a few are available, the structure can only be held together by the assistance of electrons from the A metal, and more of the latter will be needed.

The nickel arsenide structure is thus an important one in alloy theory. Though it appears to be an electron compound of electron/atom ratio 2·5, it can include a wide range of compositions, and can be related on the one hand to salt-like structures such as CdI_2 and on the other to the γ brass structure. It represents a type of transitional lattice which can accommodate itself to several requirements.

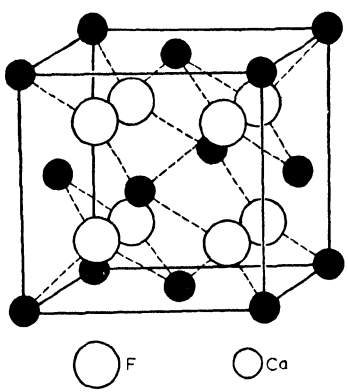


Figure 28. The calcium fluoride structure

The Fluorspar Structure—Reference has been made earlier to the compounds Mg_2Si, Mg_2Ge, Mg_2Sn, and Mg_2Pb. These all crystallize in the fluorspar structure[67] (*Figure 28*) in such a way that electropositive magnesium atoms occupy the sort of position that electronegative fluorine atoms occupy in the CaF_2 structure. Thus the structures of these compounds may be said to be 'anti-isomorphous' with calcium fluoride. The crystal structure is similar to that of an ionic crystal; Hume-Rothery has pointed out several other resemblances to ionic compounds.[1] It is therefore legitimate to consider these compounds as normal valency compounds in which the valencies of the components are satisfied. Similarly for MgS and MgSe (NaCl type of structure) and Mg_3As_2, Mg_3Sb_2, and Mg_3Bi_2 (structures anti-isomorphous with those of rare earth oxides), salt-like structures are obtained. The formation of the whole class may be considered as an extension of ordinary ionic reactions into the region of the weakly electronegative elements of Groups IVB, VB, and VIB. The electrical conductivities are low compared with those of pure metals, but the compounds do show some metallic behaviour.

An alternative point of view is taken by MOTT and JONES,[4] who show

that the first Brillouin zone for the fluorspar structure is identical in form with that for the body-centred cube, and can contain just 8/3 electrons per atom, which is just the number available in Mg_2Si and its analogues. It is therefore possible to regard these compounds as 'full zone' electron compounds of exactly the same type as the γ brass structure. The low conductivity, on this view, is due not to polar binding but to the very small number of electrons free to carry current, and it is significant that on melting, which destroys the zone structure, the conductivity of Mg_2Sn rises to approximately that of liquid tin. This could be accounted for on the ionic hypothesis by assuming complete dissociation on melting, but such behaviour among salts is not general. In an attempt to decide between these alternative views, HUME-ROTHERY and RAYNOR[68] prepared ternary alloys of magnesium and tin with small amounts of aluminium or indium, with the electron/atom ratio adjusted to 8/3. In no case was a homogeneous alloy obtained, so that if Mg_2Sn is an electron compound its characteristics are different from those of the copper, silver, and gold electron compounds. This is not conclusive evidence for ionic character, since the accommodation of foreign atoms in a complex structure like that of Mg_2Sn may be difficult.

Recent researches have discovered many compounds which crystallize in the fluorspar structure, including:

Mg_2Sn	Mg_2Pb	Mg_2Si	Mg_2Ge	Li_2S	Na_2S
Cu_2S	Cu_2Se	Be_2C	Sn_2Pt	Al_2Ca	CuCdSb
CuMgSb	CuBiMg	AgAsMg	LiMgN	LiZnN	Li_3AlN_2
		Li_3GaN_2			

All these compounds have 8/3 electrons per atom; in view of this, and of the fact that in Al_2Ca normal valencies are not satisfied, it seems probable that in these intermetallic compounds the controlling factor is the number of electrons per atom.

In a further series of fluorspar-type compounds:

$AuAl_2$ $AuGa_2$ $AuIn_2$ $PtAl_2$ $PtGa_2$ $PtIn_2$ $NiIn_2$

it must be assumed, in order to obtain the electron/atom ratio 8/3, that gold, platinum, and nickel each contribute two electrons per atom. For gold this may be understood in terms of the deformation of the relatively large gold ion by the small heavily charged aluminium ions. A similar process may occur for the transition metals. The compound

NiMgSb has a structure closely similar to that of fluorspar, and the effective valency of nickel would appear to be unity in this compound.

The two points of view from which the fluorspar structure can be regarded are not mutually exclusive. It is quite possible for electron compounds to have a marked heteropolar character. It is of great interest to note that if the electron compound interpretation be accepted, then the eight electrons can be regarded as approximately equally shared between the three atoms. Each magnesium atom has an extra 2/3 electron and is hence negatively charged, while each tin atom has a deficit of 4/3 electrons and is hence positively charged. The ratio (negative charge on Mg)/(positive charge on Sn) = 1:2, which is the same as the ratio (negative charge on F)/(positive charge on Ca) in the fluorspar structure. This completely accounts for the anti-isomorphism of the two structures.

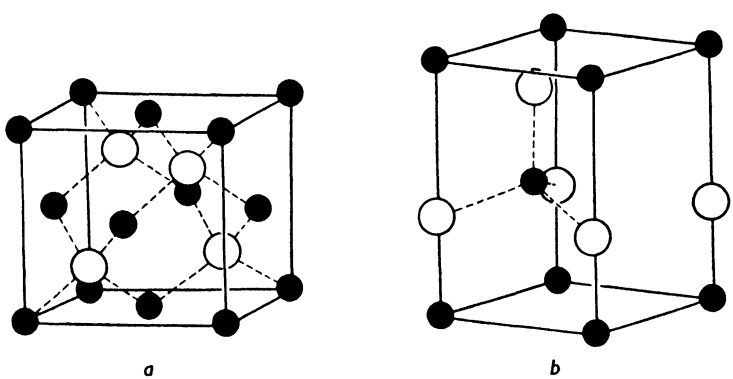


Figure 29. The structure of a zinc blende b wurtzite

The Zinc Blende and Wurtzite Structures—These structures are assumed by several AB compounds and are closely related. *Figure 29a* shows the zinc blende structure, while that of wurtzite is given in *Figure 29b*. Each atom is tetrahedrally linked to four close neighbours, the zinc blende arrangement being identical with that of diamond. These structures have been discussed by HUME-ROTHERY,[1] and it will be sufficient to note here that all known compounds assuming these structures have an electron/atom ratio of 4. They can thus be regarded as 'full zone compounds'; at the same time it must be recognized that the tetrahedral arrangement indicates covalent bonding. This class of compound may thus be interpreted in different ways.

The discussion in this section shows that much progress has been made in classifying intermediate phases in alloys in terms of electron/atom ratios, relative atomic sizes, and electrochemical effects. Attention is now being directed towards a closer study of the binding forces in the various main classes, and in this connection the new developments in the thermochemistry of alloys[69] are likely to be important. It is clear, however, that the problem must be viewed as a whole, since there are many compounds of behaviour intermediate between those of the main types. The influence of valency in the Laves phases illustrates this, and emphasizes the difficulty of making strict classifications.

Summing up, electron compounds of the β, γ, and ε brass types will only be formed, for alloys with copper, silver, gold, and the transition metals, when the electron/atom ratio is between 1 and 2 and when the ratio of the atomic radii of the components is less than 1·2. If these conditions are not satisfied, Laves phases or nickel arsenide phases may occur. A strong electrochemical difference between the components favours the formation of structures of the type of MgTe, Mg_2Sn, and Mg_3Sb_2 but may also give rise to heteropolar compounds in which the valencies of the components are not satisfied. Such structures are the result of a compromise between the tendency of ionic compounds to have a low coordination number (maximum eight) and the tendency of metallic structures to have high coordination numbers. Such compounds are likely to occur for alkali, alkaline earth, and rare earth metals with most other metals except beryllium, magnesium, and the transition metals. The transition metals tend to form nickel arsenide phases with many metals. The limiting condition for the formation of all classes of intermediate phases seems to be that the atomic radius ratio should not exceed 1·38.

THE BEHAVIOUR OF TRANSITION METALS IN ALUMINIUM ALLOYS

The theoretical considerations discussed earlier indicate that the valency of a transition metal in an alloy may vary according to the nature of its partner and the amount present. We shall now consider more fully the absorption of electrons by transition metals.

According to the theories of Mott and Jones,[4] the electronic structures of the transition metals of the first long period in the solid state are such that the 3*d* and 4*s* bands, which can contain ten and two

electrons per atom respectively, overlap so that there are vacant states at the top of the *d* band. Nickel, with ten outer electrons, has 0·6 electron per atom in the *s* band and a corresponding number of vacancies in the *d* band. Cobalt, with one electron per atom less, has about the same number of *s* electrons but a greater number of vacancies per atom in the *d* band. This model accounts well for the electrical and some of the magnetic properties of the transition metals but regards the electrons as belonging to the whole lattice and neglects their utilization in bond formation between the atoms. This problem has been considered by PAULING,[70] who has proposed a theory of the transition metals which is more revealing for cohesion problems and probably for the subject under discussion here.

According to Pauling, the 3*d* and 4*s* electrons must be considered as a whole in assessing the nature of the bonding, and in development of earlier work on the nature of the chemical bond in many classes of substance he suggests that of these electrons approximately 5·78 per atom are concerned with bonding, while the remainder occupy a band of energy levels which Pauling calls the atomic orbitals. There are 2·44 atomic orbitals, each of which can accommodate one electron of positive spin and one electron of negative spin. The electrons occupying the atomic orbitals, however, tend to remain unpaired as long as possible, *i.e.* to have no corresponding electron of opposite spin. Thus chromium, with 6 electrons outside the argon-like core, has 5·78 bonding electrons and 0·22 unpaired electrons in atomic orbitals. Manganese, with one electron more, has 1·22 unpaired electrons in atomic orbitals, while iron has 2·22 unpaired atomic orbital electrons. Since there are 2·44 atomic orbitals there is room for only 2·44 unpaired electrons, so that for cobalt, with one electron more than iron, the pairing of electrons of opposite spin must begin. For nickel the extra electron produces further pairing. The number of atomic orbitals was deduced from the saturation magnetic moment of iron and its alloys. The saturation magnetic moment of iron is 2·22 Bohr magnetons per atom, which means that there are, in solid iron, 2·22 electrons of the same spin quantum number without any corresponding electrons of opposite spin. When cobalt is dissolved in iron the number of electrons to be accommodated in atomic orbitals increases, and experimentally it is found that the first additions raise the saturation magnetic moment to a maximum of 2·44 Bohr magnetons per atom, at which stage

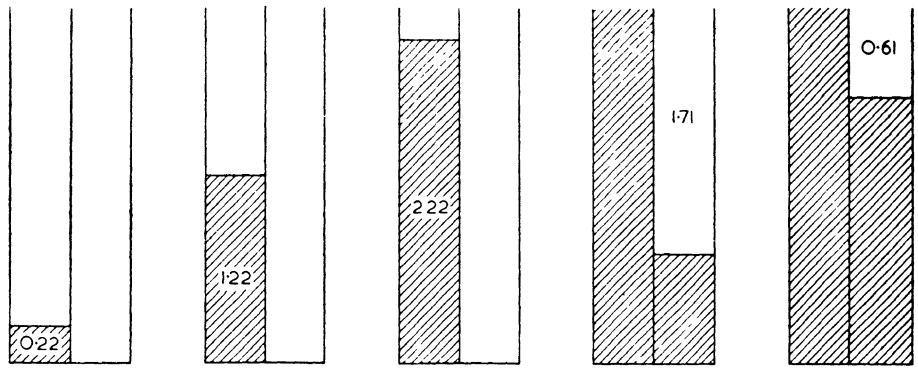
further additions cause a decrease. Thus the number of unpaired atomic orbital electrons can increase to 2·44 per atom, but further electrons must pair with those already present. The electron distribution in the transition metals is summarized in *Table VIII*. In this the atomic orbitals have been represented as divided compartments into the two halves of which electrons with opposite spin quantum numbers can be accommodated. The rigid application of the theory would lead to 1·66 and 0·66 unpaired electrons for cobalt and nickel respectively; the actual saturation moments are 1·71 and 0·61, and these have been used in *Table VIII*.

The connection between Pauling's theory and the more usual band theory is difficult to see. For the transition metals, however, Pauling's theory may be 'translated' into band theory language as follows. The 3*d* band, with room for ten electrons, is considered as divided into two sub-bands separated by an energy gap. One can contain a maximum of 4·88 electrons (*Table VIII*) and the other can contain 5·12 electrons. In addition, for all the transition elements there is approximately 0·66 electron in the 4*s* band. The 3*d* sub-band with 5·12 electrons and the 4*s* electrons are both concerned in bonding, and if this number of electrons is regarded as constant, alterations in the number of electrons per atom can only affect the 3*d* sub-band containing 4·88 electrons. The Pauling theory, accounting for the cohesion, approximate constancy of lattice spacings, and high melting points of the transition metals, may be regarded as a refinement of the band theory to take into consideration the way the atoms form bonds with each other.

Table VIII shows that chromium, manganese, iron, cobalt, and nickel have respectively 4·66, 3·66, 2·66, 1·71, and 0·61 vacancies per atom in the atomic orbitals. It appears probable that in alloys these vacancies may be filled by electrons contributed by the other component, especially where this has a high valency. According to the band theory, the vacancies per atom in the 3*d* shell which may be filled by electrons contributed by the other component would be respectively 4, 3, 2, 1, and 0 for the elements chromium to nickel. In general, experimental evidence derived from studies of alloys favours the non-integral values, which are used in this chapter. Aluminium, with three valency electrons per atom, is a suitable element for the study of these effects, which will now be considered.

The binary compounds formed by aluminium and the transition

Table VIII. Summary of Electron Distribution

Metal	Cr	Mn	Fe	Co	Ni
Configuration in free atom ..	(2) (2,6) (2,6,5) 1	(2) (2,6) (2,6,5) 2	(2) (2,6) (2,6,6) 2	(2) (2,6) (2,6,7) 2	(2) (2,6) (2,6,8) 2
Electrons in bonding orbitals ..	5·78	5·78	5·78	5·78	5·78
Unpaired electrons in atomic orbitals	0·22	1·22	2·22	1·71	0·61
Paired electrons in atomic orbitals	—	—	—	1·51	3·61
Total electrons	6	7	8	9	10
Number of vacancies in atomic orbitals	4·66	3·66	2·66	1·71	0·61
Distribution of electrons in atomic orbitals					

metals which come into equilibrium with the primary solid solution are respectively $CrAl_7$, $MnAl_6$, $FeAl_3$, Co_2Al_9, $NiAl_3(CuAl_2)$. Except for $FeAl_3$, the proportion of aluminium decreases as the number of vacancies per atom in the atomic orbitals of the transition metal decreases. This suggests a connection between the composition and the extent to which electrons may be absorbed by the transition metal atoms.[71] If the electron/atom ratios for these compounds are calculated assuming that each aluminium atom contributes three electrons to the structure while each transition metal atom absorbs electrons to the maximum extent (*i.e.* 0·61 for nickel, 1·71 for cobalt *etc*), an approximately constant electron/atom ratio is maintained, except in the case of $FeAl_3$:

$CrAl_7$	$MnAl_6$	$FeAl_3$	Co_2Al_9*	$NiAl_3$
2·05	2·05	1·58	2·12	2·09

This suggests that such compounds are a type of electron compound, whose formation is governed by the attainment of a given electron/atom ratio. Support is given to this view by the isomorphy of Co_2Al_9 and the ternary compound $FeNiAl_9$.[72] In each case there are nine aluminium atoms, and the number of vacancies in two atoms of cobalt is very closely similar to the number of vacancies in one atom of iron plus one atom of nickel; the electron/atom ratios are thus almost identical. The constitution of the aluminium-rich aluminium-iron-nickel alloys has been investigated by BRADLEY and TAYLOR[72] and independently by RAYNOR and PFEIL,[73] and the essential features are shown in *Figure 30*.

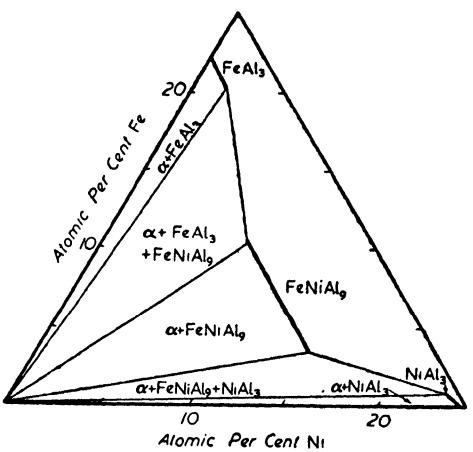


Figure 30. The 600°C isothermal of aluminium-iron-nickel alloys

$FeNiAl_9$ has a wide range of solubility for nickel, and along the boundary at which it is in equilibrium with the primary solid

*The composition of this phase contains slightly more cobalt than corresponds to Co_2Al_9. The electron/atom ratio is calculated for the experimental composition.

solution iron is replaced by nickel strictly atom for atom until a limit is reached at 2·285 electrons per atom. Since the atomic diameters of nickel and iron are similar there seems little reason for the replacement process to stop unless at 2·285 electrons per atom the Brillouin zone for the structure is full. If the electronic interpretation of the form of this diagram is correct, it implies that the form of the aluminium-rich aluminium-cobalt-nickel diagram should be analogous to that of the aluminium-nickel-iron alloys on the nickel-rich side of the line representing Fe : Ni = 1, while the form of the aluminium-rich aluminium-iron-cobalt diagram should be analogous to that of the aluminium-iron-nickel alloys on the iron-rich side of the same line, *i.e.* the aluminium-iron-nickel diagram is to be regarded as built up from the other two diagrams placed side by side.[73] The aluminium-cobalt-nickel and aluminium-iron-cobalt equilibrium diagrams have now been examined, and striking analogies exist.[74,75] In neither case is a ternary compound formed, in accordance with expectation, and there is a wide range of solubility of nickel in Co_2Al_9, corresponding to that in $FeNiAl_9$; in both cases the replacement is strictly atom for atom and proceeds to the same electron/atom ratio of 2·285. This supports the interpretation of this type of compound as controlled by electronic effects. The accuracy with which the two simpler ternary diagrams build up the aluminium-iron-nickel diagram is shown in *Figures 30* and *31*. In *Figure 31* the 600°C isothermals for each simpler diagram are plotted inside a triangle with angles of 60°, 30°, and 90°. The two placed together with the Al-Co axis common make up the usual equilateral triangle used for plotting ternary systems, and, except for a somewhat larger solubility of iron in Co_2Al_9 than in $FeNiAl_9$, the agreement with *Figure 30* is almost exact.[75] It is also noteworthy that $FeAl_3$ dissolves both nickel and cobalt to the same electron/atom ratio.

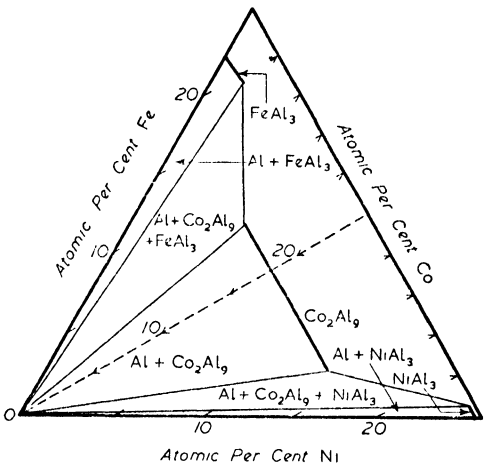


Figure 31. Comparison of the 600°C isothermals of the aluminium-cobalt-nickel and the aluminium-cobalt-iron alloys with that of the aluminium-iron-nickel alloys

Further application of these principles has enabled prediction of the form of the aluminium-rich region of the quaternary aluminium-iron-cobalt-nickel equilibrium diagram, in excellent agreement with subsequent experiment.

It will be noted that the excess of electrons which Co_2Al_9 can accommodate is much greater than the deficit it can tolerate. This is a common characteristic of electron compounds, in which a deficit of electrons is less easily tolerated than an excess; thus many γ brass phases extend considerably above 21/13 electrons per atom, but only slightly below. The same tendency is shown by the other binary compounds of aluminium and the transition metals listed on page 49. Each will dissolve much more of the transition metals which raise the electron/atom ratio, than of those metals which decrease it. The extent to which an electron deficit can be tolerated appears to depend on the precise details of the structure; the stable Co_2Al_9 can, for instance, dissolve more iron than the less stable $FeNiAl_9$ in which half the transition metal sites in the lattice are occupied by iron and half by nickel.

Other interesting analogies have been traced. In the aluminium-manganese-copper system a ternary compound exists, based on the arrangement $Cu_2Mn_3Al_{20}$ or $(CuMn)Al_4$. Also, in the aluminium-manganese-nickel system a ternary compound $Ni_4Mn_{11}Al_{60}$ occurs. According to the theory of the role of transition metals developed in this section, the electron/atom ratios of this phase, and of the $(CuMn)Al_4$ phase at the composition in equilibrium with $MnAl_6$ and the primary solid solution, are both close to 1·85.[71] For these compounds, therefore, two conditions must be satisfied: the ratio of solute atoms to aluminium atoms must be 1:4, and the electron/atom ratio 1·85. Subsequent work has confirmed the prediction of an analogous compound in the system aluminium-manganese-zinc.[76] *Figure 32* shows the homogeneity

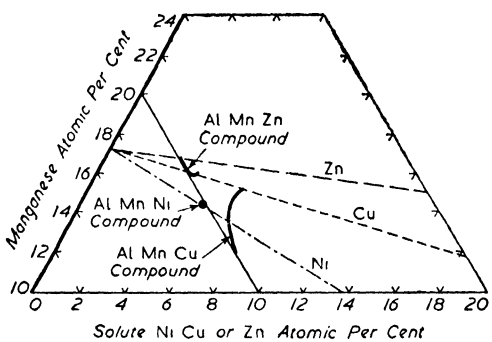


Figure 32. The homogeneity ranges for the aluminium rich ternary compounds in the systems aluminium-manganese-zinc, aluminium-manganese-nickel, and aluminium-manganese-copper. The dotted lines represent an electron/atom ratio of 1·85 for each metal

ranges for these three compounds, and it is seen that they are grouped round the intersection of lines representing the correct atomic ratio, and the electron/atom ratio 1·85 for the various alloy systems. The attainment of the correct atomic ratio will depend on atomic size relationships. Thus, when zinc is replaced by cadmium, which has both a larger atom and a larger ion, no such compound is expected or found.[77] Also, when copper is replaced by silver, similar arguments again suggest no ternary compound of this type will be found; this has also been confirmed.[78] There would appear, however, to be little reason why an analogous compound should not be found in the aluminium-manganese-magnesium system, since although the atomic size of magnesium is greater than that of zinc, the ionic diameter is approximately the same. Although LEEMANN[79] reported no ternary compound in this system, a re-examination[78] has shown that a compound does come into equilibrium with the primary solid solution. Its electron/atom ratio, however, is not 1·85, and the compound appears to be analogous to the ternary compound which occurs in the aluminium-chromium-magnesium alloys.

There appears to be a further class of compound in which no definite ratio of solute atoms to aluminium atoms is maintained, though the electron/atom ratio is again approximately constant. Thus in the systems aluminium-iron-copper and aluminium-nickel-copper, ternary compounds which come into equilibrium with the primary solid solution occur at $FeCu_2Al_7$ and Cu_3NiAl_6. If the valency of iron and cobalt were zero, the respective electron/atom ratios would be 2·3 and 2·1. If, however, the ideas discussed above are applied, the electron/atom ratio would be 2·04 in each case. This interpretation would imply that an analogous ternary compound of composition $Co_2Cu_5Al_{13}$ should be found in the aluminium-cobalt-copper system, and this has been confirmed.[80]

Further indications of the importance of the electron/atom ratio are given by the recent proof[81] that the complex ternary compounds in the aluminium-manganese-silicon and aluminium-iron-silicon systems, usually denoted respectively as α(MnSi) and α(FeSi), which are the end members of a continuous series of solid solutions in the quaternary aluminium-manganese-iron-silicon system, have the same electron/atom ratio. It is further significant that the compound Cr_2Al_{11} will dissolve manganese in such a way that almost the whole of the chromium

is replaced by manganese, the electron/atom ratio remaining unchanged throughout.[82]

The recent determination of the crystal structure of the compound Co_2Al_9[110] has given further support to the conception of the absorption of electrons by transition metal atoms in aluminium alloys. Electron density maps for this structure, constructed from the crystallographic data, indicate that the cobalt atoms are each associated with twenty nine electrons instead of their normal complement of twenty seven.[110] This estimate was made on the assumption that the free electron density between the atoms was zero, which may not be strictly correct; however, the density of free electrons in the region of a cobalt atom which has a negative charge by reason of its absorption of electrons may be expected to be low, so that the evidence in favour of absorption is strong. It has also been shown that the Brillouin zone for Co_2Al_9, deduced from the new crystallographic data, is of such a form that the inscribed sphere corresponds to almost exactly the same number of electrons per atom as that calculated for the experimentally determined composition of Co_2Al_9, assuming that cobalt absorbs 1·71 electrons per atom.[110–112] Similarly, the form of the Brillouin zones for $NiAl_3$ and Co_2Al_5, the crystal structures of which are known, are also consistent with the general theory.

A further large class of alloy systems and intermetallic compounds has thus been brought into line with alloy theory, and it is clear that, when alloyed with metals of relatively high valency, transition metal atoms will absorb electrons from the structure as a whole. The experimental evidence suggests that the extent to which electron absorption takes place is very similar to the number of vacancies per atom in the atomic orbitals of the transition metals postulated by PAULING. The physical mechanism of the process on this basis is, however, somewhat difficult to understand. It is possible that, with the accumulation of data, evidence for the absorption of integral numbers of electrons as suggested by the band theory may become stronger; this would have the advantage of providing an easily understood physical mechanism of absorption into the incomplete 3*d* band of the transition metal. The structures of the binary and ternary compounds of aluminium with transition metals appear to be of the electron compound type, but, as in all instances, other factors, such as the size factor, are involved. The theory may be applied to the prediction of unknown or partially known phase diagrams, and a very interesting stage has been reached.

TERNARY EQUILIBRIUM

In recent years attention has been paid to the factors affecting the formation of ternary alloys and the nature of the equilibrium involved. Some of the principles will be apparent from other sections; further developments are discussed here.

It has been shown by Hume-Rothery that in a ternary system ABC, where B and C form a stable compound B_xC_y of substantially fixed composition, the form of the solid solubility and liquidus isothermals of B and C in A depends on the nature of the compound.[12] The phenomenon may be interpreted in terms of a modified solubility product law:

$$[B]^x[C]^y = \text{constant } K$$

The square brackets refer to atomic percentages corresponding to points on an isothermal, *i.e.* a liquidus isothermal or the $\alpha/(\alpha + B_xC_y)$ isothermal. The rigorous expression would include volume concentrations and activity coefficients, but reasonable accuracy is obtained by the simple equation. Thus, for the precipitation of Mg_2Si from the solid solution of magnesium and silicon in aluminium, we may write:

$$[Mg]^2[Si] = \text{constant } K, \text{ or } 2 \log [Mg] + \log [Si] = \log K$$

In agreement with this, liquidus and solid solubility isothermals give a straight line of the correct slope (except at low concentrations of one of the components) when log [Mg] is plotted against log [Si] for a series of points along the appropriate phase boundary.[12] Different temperatures correspond to different values of K, and, if the constants for a series of temperatures are evaluated, it is found that a relation $\mathrm{d} \log K/\mathrm{d}T = -Q/RT^2$ is obeyed, where T is the absolute temperature and Q is a constant related to the heat of formation of the compound. In principle, if sufficient data are available to show that a given system ABC obeys these laws, the whole of the liquidus surface at which B_xC_y is primary, and the whole of the $\alpha/(\alpha + B_xC_y)$ boundary, can be calculated.

Similar relationships hold in some cases where the precipitating phase is a ternary compound. Thus in the system aluminium-magnesium-zinc a ternary compound $Mg_3Zn_3Al_2$ comes into equilibrium with the primary solid solution, and the $\alpha/(\alpha + Mg_3Zn_3Al_2)$ isothermals may be expressed by the relation:[83]

$$[Mg]\,[Zn] = K, \text{ where } \mathrm{d} \log K/\mathrm{d}T = -Q/RT^2$$

The compound, however, has a relatively wide range of homogeneity, and the existence of this relation was surprising. It was shown, however, that the portion of the boundary over which the relation holds accurately

represents equilibrium with $Mg_3Zn_3Al_2$ of such a composition that magnesium and zinc are present in equiatomic proportions.[84] Other systems in which similar laws apply are the systems copper-magnesium-silicon (compound Mg_2Si) and copper-nickel-tin (compound Ni_3Sn_2).

In the absence of compound formation, solid solubility isothermals in copper, silver, and gold are controlled mainly by electronic and size effects.[12] Thus for copper-aluminium-zinc alloys, where both solutes have favourable size factors, the isothermals are straight lines joining the limits in the binary systems and have the form of *Figure 33a*.

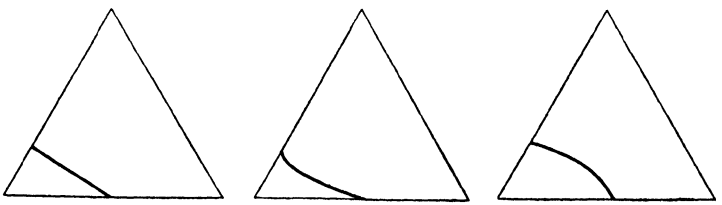


Figure 33. Solid solubility isothermals a ideal b with accentuated lattice distortion c when one solute relieves the distortion due to the other

Increasing distortion of the solvent lattice tends to depress the electron/atom ratio corresponding to maximum solubility, so that if one solute tends to accentuate the distortion produced by the other, as in the copper-aluminium-tin system, deviations from linearity occur. Tin atoms cause a relatively intense local distortion of the copper lattice, so that as aluminium is replaced by tin the electron/atom ratio at saturation tends to fall. Also, since the lattice is already strained by a large number of aluminium atoms, the distortion round the introduced tin atoms cannot readily be taken up by the lattice. The solubility is thus lower than in the ideal case. As the replacement process continues, the total number of solute atoms decreases. Therefore it becomes easier for the distortion round the tin atoms to be taken up, because the total number of centres of distortion is decreasing, and at a certain stage the restriction of the solubility will decrease. The resultant curve appears as in *Figure 33b*, and is also shown by copper-zinc-tin and copper-zinc-cadmium alloys. If one solute relieves the distortion due to the other the curve of *Figure 33c* will result; this behaviour is shown by zinc-rich zinc-aluminium-copper alloys ($D_{Al} > D_{Zn} > D_{Cu}$) and silver-rich silver-antimony-zinc alloys ($D_{Sb} > D_{Ag} > D_{Zn}$), where D represents the atomic diameter.[12]

Some principles governing intermediate phases in ternary systems have already been discussed. Another general principle is that electron compounds of the 3/2, 21/13, and 7/4 types tend to form complete series of solid solutions with others of the same electron/atom ratio. In copper-zinc-aluminium alloys the γ brass structure is maintained if the proportions of the components are adjusted to keep the electron/atom ratio at 21/13.[85] Similarly the γ phases of the systems silver-zinc and copper-zinc form a continuous series of solid solutions.[86] Continuous series of solid solutions are also formed, at the appropriate temperatures, between the body-centred cubic β phases of the systems copper-zinc-tin, copper-zinc-aluminium, copper-aluminium-tin, and other alloys. The low temperature forms of Ag_3Ga and AgZn, both of which are close-packed hexagonal, form similar unbroken solid solutions,[87] while the hexagonal Ag_5Sn and AgCd phases show the same behaviour.[88] Ag_5Sn and $AgCd_3$ (a hexagonal 7/4 electron compound) also form extensive solid solutions, but there is a small miscibility gap. Among the 7/4 electron compounds continuous solid solutions are formed between:

$AgZn_3$ and Ag_5Al_3 [89]	$AgZn_3$ and $CuZn_3$ [86]
εMnZn and $AgZn_3$ [86]	εMnZn and $CuZn_3$ [86]

while in the quaternary system silver-copper-manganese-zinc the 7/4 electron compounds form a complete quaternary series of solutions. Similar electron compounds thus tend to be mutually soluble. The cubic CuBe and the 3/2 phase Cu_5Si with the β manganese structure do not, however, dissolve mutually to any extent.[90]

Within such continuous ranges of solid solution the effects of size factors, electronic and other influences can be traced.[91] The copper-zinc β phase is stable to room temperature, while that of the copper-tin system decomposes eutectoidally at 586°C. In ternary copper-zinc-tin alloys, the first additions of tin to the copper-zinc phase do not alter the form of the equilibrium diagram. Further additions introduce a eutectoid reaction, and with 10 per cent of tin the form of the diagram is strikingly similar to that of the copper-aluminium alloys (*Figure 34*). This behaviour is due to the variation in atomic size relationships from the copper-zinc to the copper-tin axes of the ternary model. It is difficult to assess size factors in ternary systems, but if it be assumed as a first approximation that the effects are additive, the effective size factor for a ternary alloy with x atomic per cent of solute B and y atomic

per cent of solute C is $(xF_B + yF_C)/(x + y)$, where F_B and F_C are the percentage size factors of B and C with respect to the solvent A. *Figure 35* shows the eutectoid temperature in the ternary system plotted

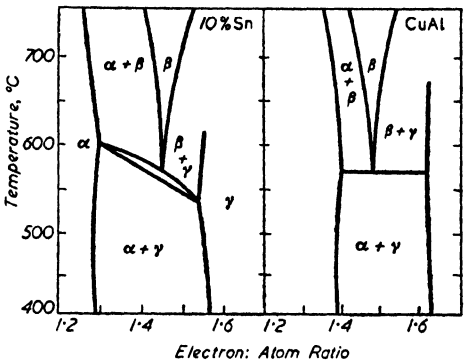


Figure 34. Comparison of the equilibrium diagrams of the ternary alloy copper-zinc-tin with 10 per cent tin and copper-aluminium alloys

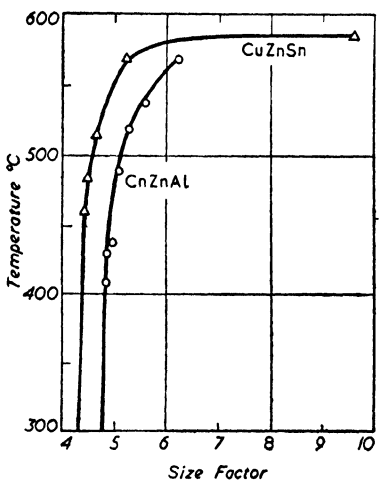


Figure 35. The relationship between the eutectoid temperature and effective size factor for ternary systems

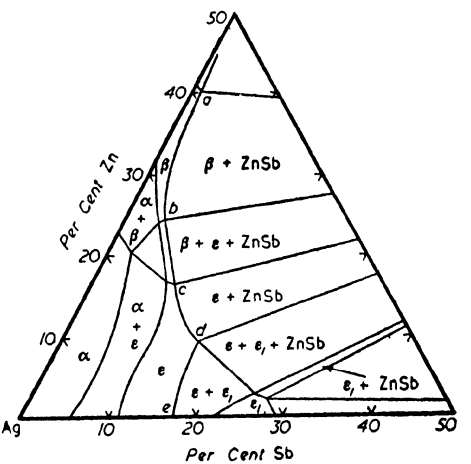


Figure 36. The 400°C isothermal of the system silver-zinc-antimony

against the effective size factor; the eutectoid temperature rises to that for binary copper-aluminium alloys at approximately the same size factor as characterizes the binary copper-aluminium alloys, *i.e.* + 6·6. This confirms that the temperature range of the β phases is sensitive to the size factor, and suggests that details of a ternary diagram ABC may be predictable from a knowledge of binary alloys, in solvent A,

of other elements where size factors lie between those of B and C with respect to A. The data for the copper-zinc-aluminium alloys also included in *Figure 35* show an analogous trend; the difference between the two curves is due to a valency factor.[91]

The 3/2 electron compounds in the system silver-zinc-antimony, the 400°C isothermal for which is shown in *Figure 36*,[92] are also of interest.[93]

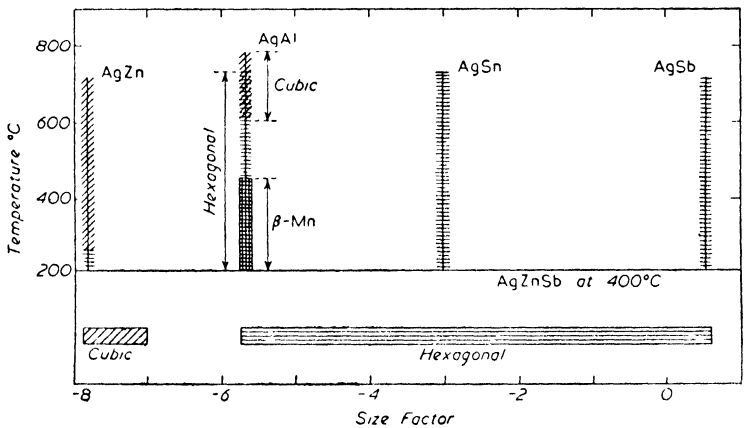


Figure 37. The relation between the 400°C homogeneity ranges of the various structures in the silver-zinc-antimony system with effective size factor

The cubic AgZn phase dissolves much antimony, and the hexagonal Ag_7Sb phase dissolves a considerable amount of zinc. The homogeneity ranges of the respective structures at 400°C are plotted against the effective size factor in *Figure 37*, and data are included for binary alloys of silver where the size factors fall within the same range. The hexagonal structure persists down to the same size factor at which it begins to be unstable relative to the cubic structure in the binary alloys. The cubic structure becomes unstable before the size factor at which the hexagonal phases occur in the binary systems is reached.

It may be noted that in *Figure 36* the boundary *cd*, where equilibrium with the compound ZnSb is involved, obeys the relation [Zn] [Sb] = K in just the same way as primary solubility isothermals. The boundaries *ab* and *de* correspond to almost constant electron/atom ratios of 1·5 and 1·62 respectively.

Having elucidated some of the details of ternary diagrams, attention has been given to the factors which determine which of the possible

phases come into equilibrium with each other. In binary systems equilibrium is determined by the points of contact of a common tangent to two free energy curves (*Figure 1*). In ternary alloys the free energy curves become surfaces, and equilibrium is determined by the positions of tangent planes which make contact with three surfaces. The condition for equilibrium is that all compositions should possess the lowest

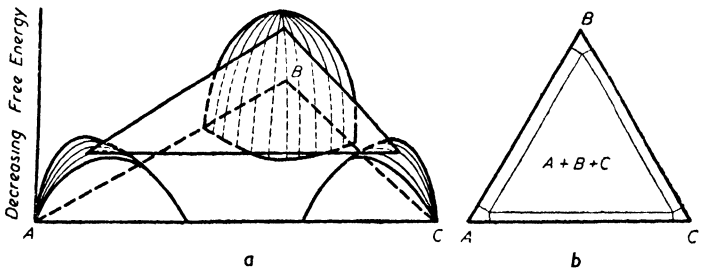


Figure 38. a Hypothetical free energy surfaces for a ternary system with no compound formation, b plan of the tangent plane at a given temperature

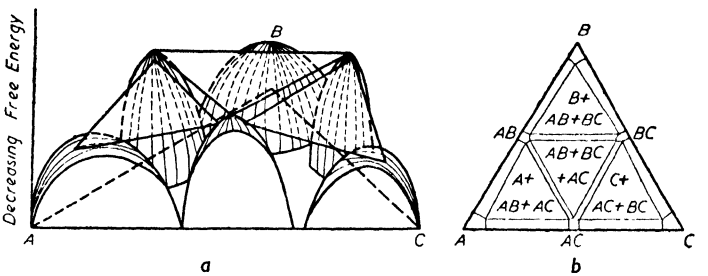


Figure 39. a Hypothetical free energy surfaces for a ternary system with the formation of three binary compounds, b plan of the tangent planes at a given temperature

possible free energy. *Figure 38a* shows hypothetical free energy surfaces for a system ABC in which there are no compounds; the surfaces are represented as 'humps', with energy decreasing upwards rather than as pits, for ease in drawing. The tangent plane determines the equilibrium, which is shown, for a given temperature, in *Figure 38b*. If now binary compounds AB, BC, and CA are formed, with approximately the same free energies, the free energy relations and equilibria will be as in *Figure 39a* and *b*. If, however, the compound BC is very stable, with a very low free energy minimum, the arrangement of tangent planes in *Figure 39* would not be appropriate; a lower free energy for much of the composition range is obtained by the arrangement of *Figure 40*. The effect is to bring BC into equilibrium with the A-rich phase.

The equilibrium obtained in a given case thus depends on the relative free energies of the possible intermediate phases. If they are comparable, equilibria of the type of *Figure 39* may be expected, but if one phase has a very low free energy minimum it will dominate the equilibria. If the free energies of the possible phases were known it would be possible to predict equilibrium relations approximately. Exact values are not

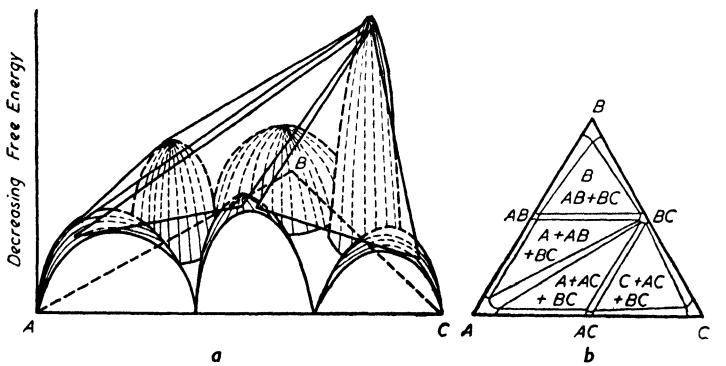


Figure 40. a Hypothetical free energy surfaces for a ternary system with the formation of three binary compounds, one of which is extremely stable; b plan of the tangent planes at a given temperature

in general available, but to a first approximation the heat of formation may be taken as a measure of the free energy. Phases with high heats of formation will have low free energy values.

Considering systems ABC, in which it is required that a compound between B and C should come into equilibrium with the A-rich solid solution, the solutes B and C must be chosen so that there is a phase in the BC system with a much higher heat of formation than any which occur in the AB or AC systems. In this connection the following systems are of interest.

Copper-Aluminium-Nickel—Copper and nickel form no compounds, while the maximum heat of formation in the copper-aluminium system is 5·35 kcal/gm atom for Cu_9Al_4. Several compounds occur in the nickel-aluminium system, and the maximum heat of formation is 17·0 kcal/gm atom for NiAl, which, in agreement with the above ideas, enters into equilibrium with the copper-rich solution.

Copper-Silicon Alloys with Cobalt and Nickel—Although no data exist for heats of formation of copper-silicon alloys, they are not likely to be high; the heats of formation in the cobalt-silicon and nickel-silicon

series, however, reach maxima of 12·0 and 11·1 kcal/gm atom respectively, and binary compounds of the transition metal and silicon precipitate from the copper-rich solution.

Copper-Nickel-Tin—Heats of formation in the copper-tin system do not exceed 1·8 kcal/gm atom, whereas the heat of formation of Ni_3Sn_2 is 7·5 kcal/gm atom; Ni_3Sn_2 comes into equilibrium with the copper-rich solution.

Copper-Tin-Magnesium—Though not very high, the heat of formation of Mg_2Si (5·6 kcal/gm atom) is appreciably higher than the maximum (1·8 kcal/gm atom) in the copper-tin system, and Mg_2Sn comes into equilibrium with the primary copper-rich alloys.

Silver-Zinc-Antimony and Silver-Cadmium-Antimony—In the systems silver-zinc and silver-cadmium the heats of formation are low (maxima of 1·8 and 1·46 kcal/gm atom respectively). Though not accurately known, heats of formation in the silver-antimony alloys are probably comparable. The heats of formation in the zinc-antimony and cadmium-antimony systems rise to maxima, for ZnSb and CdSb, of approximately 1·8 kcal/gm atom. There are thus no phases of outstandingly high heat formation, and neither ZnSb nor CdSb come into equilibrium with the silver-rich solid solution.

Thus, for copper and silver as solvents, compounds between the solutes may come into equilibrium with the solvent-rich alloys if their heats of formation are above 5 to 6 kcal/gm atom. A very stable compound may even prevent solid solution formation between similar electron compounds.

THE LATTICE SPACINGS OF ALLOYS

The substitution of a foreign atom into a metal lattice causes local distortion, and though the interatomic distances in the bulk of the material remain unchanged, those in the vicinity of the foreign atom change. The mean interatomic distance is thus different from that of the solvent and this mean distortion may be determined by x-rays. It may be taken as a measure of the local distortion but does not give the actual magnitude of the local effects.

If the distortion effects depended only on size differences between the component atoms, the lattice spacing of any alloy should be calculable from those of the pure metals. This 'additivity law' was proposed by VEGARD.[94] Since then, much work has been carried out on

lattice spacings, and it is clear that Vegard's law is the exception rather than the rule. The reasons for deviations from linearity are of great interest in the theory of alloys, and much is gained from systematic study. For instance, if there is a tendency to form compounds *i.e.* the electrochemical factor is high, the observed lattice constants will be less than those calculated additively, owing to interaction between the atoms.[95] Repulsive interactions may also arise, giving deviations in the opposite sense, as in the copper-gold alloys.

Alloys of Copper, Silver, and Gold—These alloys are particularly favourable for the investigation of valency effects, since the number of electrons per atom can be increased considerably before any abnormal energy changes occur as shown earlier, and since size effects can be reduced to a minimum by choosing as solute metals the B sub-group elements immediately following the solvent metals in their respective periods of the periodic table. A systematic examination has been carried out by HUME-ROTHERY, LEWIN, and REYNOLDS.[96] For binary alloys of silver with cadmium, indium, tin, and antimony, increasing solute valency produces an increased lattice distortion; in dilute solution, where the lattice-spacing/composition curves are linear, the distortions produced per atom by cadmium, indium, tin, and antimony are in the ratio 2 : 3 : 4 : 6. For copper alloys with zinc, gallium, and germanium an analogous relation exists, in which the distortions produced by one atomic per cent of the solutes in the order given are in the ratio 3 : 4 : 4·8. The general conclusions of this work were confirmed by OWEN and ROBERTS,[97] who consider that the validity of exact whole-number relationships is not completely established. The valency of the solute clearly has a dominating effect on the distortion produced, but the effect is not strictly proportional to valency.

From the relative distortions produced by the solutes of different valency in copper and silver it is seen that the effect of change of valency is less marked for the former; the distortions for zinc and gallium in copper vary as 3 : 4, while those for cadmium and indium in silver vary as 2 : 3. Although increasing valency tends to expand the lattice, this tendency appears to be opposed by some factor which is more effective in the copper than in the silver series. Hume-Rothery, Lewin and Reynolds suggest that the opposing factor may be due to the gradual closing up of the electron 'shells' of the ions with increasing atomic number, which would be expected to be more important for the

copper series since the change in atomic number between each successive element is a greater proportion of the atomic number in the copper series than in the silver series.

Further interesting relationships for binary copper, silver, and gold alloys have been reported by OWEN.[16] *Figure 41* shows the distortion per atomic per cent plotted against valency for solutes of the first long period, while *Figure 42* gives similar data for solutes of the second long period. The linearity of the curves for zinc, gallium, and germanium in silver, copper, and gold is apparent. Arsenic

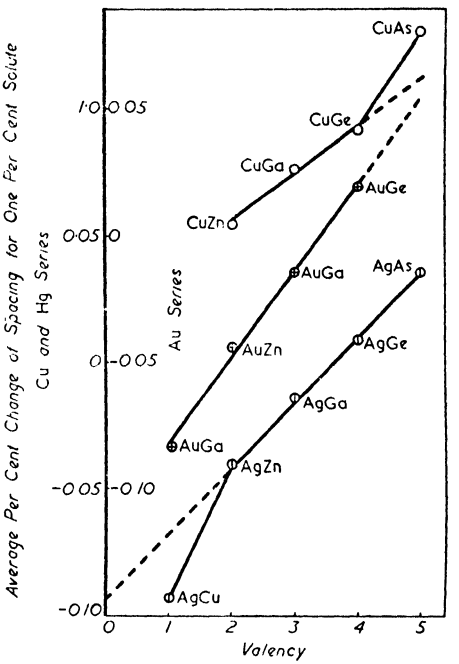


Figure 41. The relation between the distortion per atomic per cent and valency for solutes of the first long period

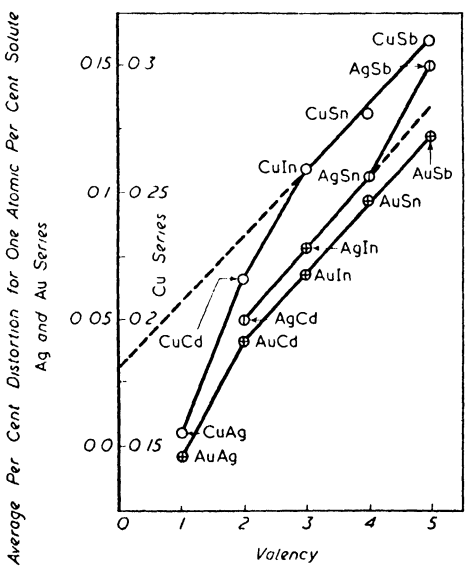


Figure 42. The relationship between the distortion per atomic per cent and valency for solutes of the second long period

behaves normally in silver but in copper behaves as if it had a valency of six. Cadmium, indium, and tin give a linear relation in silver, but, like arsenic in copper, antimony gives an abnormally large expansion. Abnormalities for copper in silver, cadmium and silver in copper, and silver in gold are apparent. It is interesting to note that the linear increase, expressed in the diagram as the percentage increase for one atomic per cent of solute, for indium, tin, and antimony in

copper is 0·0252 per valency electron, whereas when indium replaces cadmium the increase is (0·0252 + 0·0178), and when cadmium replaces silver the corresponding change is (0·0252 + 2 × 0·0178). The reason for this is obscure. Further work is also necessary to explain why antimony behaves irregularly in silver but not in copper or gold, and why arsenic behaves irregularly in copper but not in silver. Owen suggests that the abnormalities for silver in gold, and copper in silver, may be

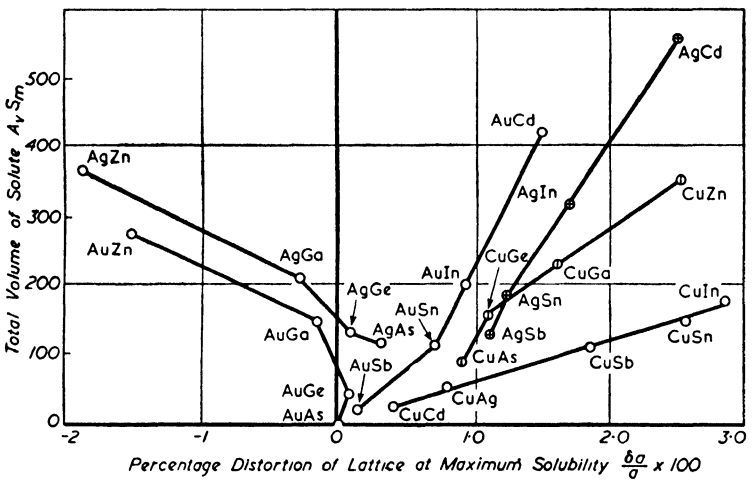


Figure 43. The relationship between the percentage lattice distortion and the total volume of solute entering the lattice at maximum solid solution

explained by a contribution by the 3*d* and 4*d* electrons of copper and silver to the cohesive forces. Displacement of one or more of the 4*d* electrons of silver would have the effect of decreasing its size. In general, to a first approximation, the size factor appears to be of little importance in this work since solutes whose size factors are both favourable and unfavourable with respect to the appropriate solvents satisfy the linear relationships in *Figures 41* and *42*.

Lattice spacing work has also shown that a relationship exists between the percentage lattice distortion, *i.e.* $(\delta a \times 100)/a$, where a is the lattice spacing, and the total volume entering the lattice at maximum solubility. *Figure 43* summarizes the available data.[16] The points for di-, tri-, and tetravalent solutes tend to fall approximately on straight lines, except for the first long-period elements in gold and silver, where the lattice spacing decreases. The range of solid solution and the total distortion are greatest for divalent solutes, except for silver and cadmium

in copper. Where the solute expands the lattice, the total volume of solute entering the lattice and the percentage distortion produced are linearly related (exceptions being found for the systems copper-arsenic, silver-antimony, and gold-antimony). Neglecting the copper-arsenic system, the straight lines for the copper alloys extrapolate to zero; the line for cadmium, indium, and tin in silver, however, does not. We may write, therefore:

$$\frac{\delta a}{a} = bv(S_m + c)$$

where $\delta a/a$ is the proportional lattice expansion, v is the atomic volume, S_m is the maximum solubility in atomic per cent, and b and c are constants which have different values for the different series of alloys. For copper alloys $c = 0$, and for silver alloys $c \simeq 10$.[97]

The effect of valency in determining the distortion in this type of alloy is clearly established, and it should be noted that a lattice expansion is observed in certain alloys where the atomic diameter, D, of the solute is smaller than that of the solvent. Thus germanium ($D = 2{\cdot}445$ Å) expands the lattice of copper ($D = 2{\cdot}551$ Å), and the case of tin in silver is similar. The distortion observed may thus be regarded as the resultant effect of the distortion produced by substituting a different sized atom for one of the solvent and that produced purely on account of the extra valency of the solute. Taking the closest distance of approach of the atoms in the metal crystals (where these are not affected by incomplete ionization) as a measure of the size of atoms, then the effective interatomic distance due to size effects only in an alloy of copper, silver, or gold (face-centred cubic structure) containing x atomic per cent solute will be $[D_{\text{solvent}} - x(D_{\text{solvent}} - D_{\text{solute}})/100] = D_A$, and the lattice spacing will be $D_A\sqrt{2}$. The observed spacings, however, are in general larger than this. The increment Δ (equal to observed spacing minus calculated spacing) has been plotted for several systems against the electron/atom ratio in *Figure 44*,[98] and it is seen that the curves for zinc in copper, cadmium in silver, and mercury in gold almost superpose, while those for germanium in copper and tin in silver similarly almost superpose. More detailed analysis shows that the percentage valency expansions of the lattice of silver by cadmium and copper by zinc per valency electron are identical; the term valency expansion is used to denote the increment Δ. Curves for gallium in

copper and indium in silver cannot be calculated, owing to doubt as to what should be taken as the closest distance to approach in the complex gallium crystal structure and the approximate nature of the corrections to be made for incomplete ionization in indium. The initial distortions produced per unit of electron/atom ratio by the extra valency of the solute atom are, for divalent and tetravalent solutes respectively, in the

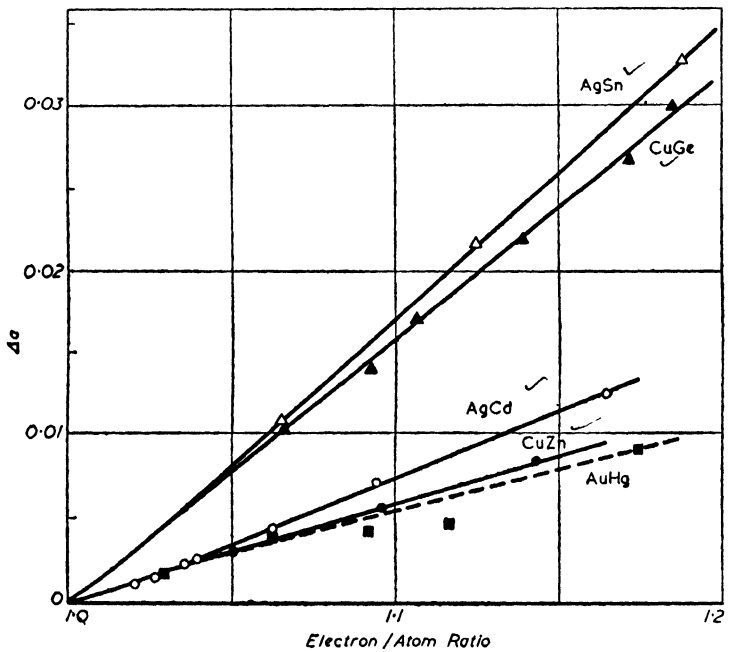


Figure 44. The relationship between the difference of observed and calculated lattice spacing, and the electron/atom ratio

ratio 1:3, suggesting that the effect is proportional to $(N - 1)$, where N is the solute valency. If this be accepted, effective atomic diameters for gallium and fully ionized indium may be calculated as 2·60 and 2·93 Å respectively. It should be noted that in *Figure 44* there is no distinction between the distortions due to valency produced by di- and tetravalent elements in copper and silver, so that the differences between the valency effects in the two solvents mentioned by Hume-Rothery, Lewin and Reynolds disappear. It seems probable that the relationships in *Figure 44* are more fundamental than the original whole-number relationships, which express the resultant effect of size differences and valency differences.

The physical mechanism by which increasing solute valency expands

the lattice is not clear. It is unlikely that the effect is due merely to the increase in the average energy of the valency electrons as their number increases. Other factors, such as repulsion between ions and the increased charge in the vicinity of the introduced polyvalent ion, are also important. For example, it cannot be assumed that the extra valency electrons are equally shared between all the atoms present. There is considerable evidence[99,100] that most of the extra electrons remain, on the average, near to the polyvalent ion; the environment of the zinc ion is, however, positive with respect to that of a copper ion in copper-zinc alloys; the environment of a copper ion is correspondingly negative compared with that of a copper ion in pure copper. It is probable that this negative charge accounts for much of the lattice expansion,[101] but the theory has yet to be properly worked out.

Alloys of Magnesium—The crystal structure of magnesium is close-packed hexagonal, and the corresponding Brillouin zone structure has been described earlier. At two electrons per atom, overlap of electrons from the first Brillouin zone occurs only in directions at right angles to the hexagonal axis and not in directions parallel to the hexagonal axis. The presence of zone overlaps, which may be modified by changing the electron/atom ratio of the structure, exerts an important influence on the lattice distortion produced by solute metals of valency other than two. It has been shown by JONES[102] that the effect of a small overlap across the bounding surface of a Brillouin zone is to tend to compress the zone in a direction at right angles to the surface concerned. Since Brillouin zones are constructed in 'reciprocal space' ($k = 1/\lambda$), this corresponds to a tendency to expand the crystal lattice in the same direction.

According to Jones' calculations, σ, the stress produced by n electrons lying within a small region of k space beyond a plane of energy discontinuity, is given by the expression $\sigma = (2n/\tau)\bar{E}$, where τ is the atomic volume and $\bar{E}$ is the mean value of the energy. The effect is thus proportional to the number of overlapping electrons, and may be evaluated quantitatively in certain cases. σ acts perpendicularly to the plane concerned.

Accurate measurements of the lattice distortions produced in magnesium by the solutes silver, cadmium, indium,[103] tin, lead, gallium, aluminium, and thallium[104] have been made, and it is clear that the zone structure of magnesium is important. The variation with

composition of the *a* spacings may be considered first. The substitution of cadmium for magnesium leaves the electron/atom ratio unaltered, and the overlaps are comparatively little affected. We may therefore take the spacing/composition curve for cadmium in magnesium as a standard curve, exhibiting the behaviour to be expected in the absence of valency changes. For silver, indium, and tin the size factors with regard to magnesium are favourable, but the spacing/composition curves vary considerably. Silver reduces the electron/atom ratio of magnesium, and the overlap at right angles to the hexagonal axis is also reduced. The internal stress produced in magnesium by this overlap is thus decreased, and the zone tends to expand and the lattice to contract. In agreement with this, silver reduces the *a* spacing of magnesium more, per atomic per cent, than cadmium, and more than would be expected from size effects alone. Indium, on the other hand, increases the overlap at right angles to the hexagonal axis, thus tending to compress the zone and to expand the lattice; the lattice spacing curve shows a marked expansion compared with the curve for the system magnesium-cadmium. These results are summarized in *Figure 45*. The curves of the *c* spacing for silver and cadmium in magnesium are likewise smooth and continuous, but that for indium in magnesium has a more complicated shape. With additions of indium less than 0·75 atomic per cent the *c* spacing decreases. Above this concentration, however, an increase takes place. Similar behaviour is shown by thallium, and though the *c* spacing curves for aluminium and gallium never rise, there are sharp changes of direction at the same solute percentage (see *Figure 46*). This behaviour may be interpreted as due to the onset, at an electron/atom ratio of 2·0075, of the overlap across the faces of the Brillouin zone which are perpendicular to the

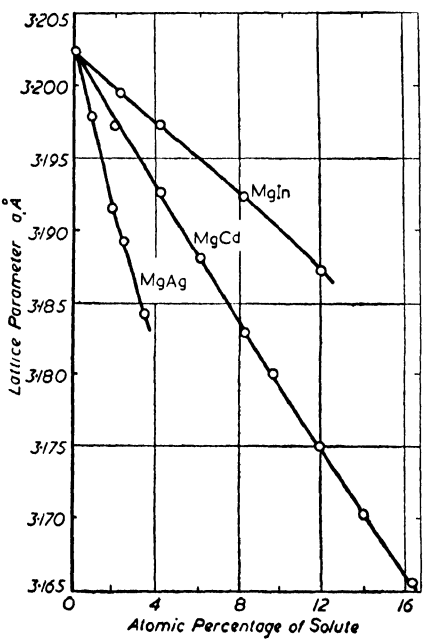


Figure 45. Lattice spacing curves for alloys of magnesium

c axis. Until this overlap occurs the *c* spacing will be relatively little affected, but once it has set in the zone is contracted and the lattice expanded in the *c* direction, so that the axial ratio *c/a* begins to rise. Experiments with tin and lead in magnesium confirm that a sharp change in direction of the *c* spacing occurs at the electron/atom ratio 2·0075. Magnesium is thus in a most interesting electronic condition; at two electrons per atom only the overlap at right angles to the *c* axis occurs, but with the addition of less than one electron per 100 atoms the second overlap occurs, causing the anomaly in the *c* spacing curves. This is confirmed by the rate of change of axial ratio for tri- and tetravalent solutes, at concentrations higher than that corresponding to the onset of the second overlap. If the rate of change of the *c* spacing is governed by the number of overlapping electrons it should be approximately twice as great, per atomic per cent, for a tetravalent as for a trivalent solute, since an atom of the former increases the number of electrons by two, as compared with one for the latter. The rate of change per unit increase in electron/atom ratio, however, should be the same in both cases.

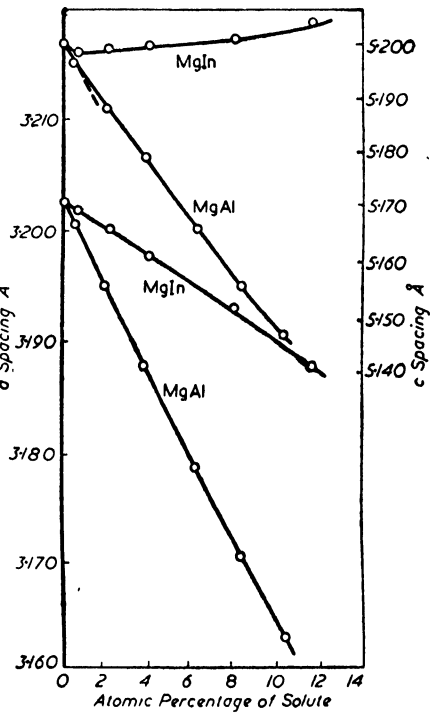


Figure 46. Lattice spacing curves for alloys of magnesium

As shown in *Figure 47*, the axial ratios for the magnesium-indium and magnesium-tin alloys superimpose when plotted in terms of electron/atom ratio. Similar superimposition is shown for magnesium-thallium and magnesium-lead alloys. Agreement between the curves for solutes of different periods is not quite exact.

Whereas the physical mechanism of the effect of valency in expanding the lattice of monovalent metals, where the electron distribution for dilute solutions is entirely within the first Brillouin zone, is not fully elucidated, in the magnesium alloys the controlling factor is the increase

or decrease in the internal stress present in the lattice, obtained by altering the number of electrons which overlap from the first Brillouin zone. This effect is, of course, superimposed on effects due to differences in atomic size.

Similar lattice spacing relationships may be expected for other structures which give rise to asymmetrical Brillouin zones, in which

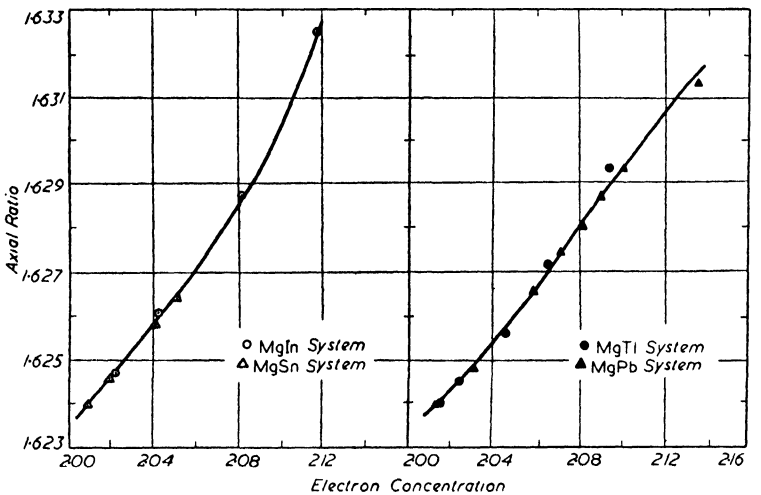


Figure 47. The relationship between axial ratio and electron concentration for magnesium alloys

some of the bounding planes are closer than others to the origin of k space. Overlap of electrons may then be expected to occur across different sets of planes at different electron/atom ratios and when a given overlap first occurs an anomaly in the appropriate lattice spacing curve may be expected. It should not be assumed that lattice spacing curves are smooth and continuous.

Similar considerations apply to the close-packed hexagonal ε and η phases in the copper-zinc system.[102] For the ε phase the axial ratio is 1·57 at the copper-rich boundary, and there is, as for magnesium, overlap of electrons from the first Brillouin zone at right angles to the hexagonal axis. On dissolving more zinc in the structure the overlap increases, and the zone tends to shrink in directions at right angles to the hexagonal axis. Thus the a spacing increases, but since the c spacing is unaffected by considerations of overlap, c/a decreases to 1·554 at the zinc-rich boundary. In the case of the η phase, the solid solution of copper in zinc, the number of electrons per atom is sufficient

to cause overlap to occur, in the saturated solution, in directions both parallel and perpendicular to the hexagonal axis, from the Brillouin zone corresponding to axial ratio 1·804. Increasing the electron/atom ratio thus affects the *c* spacing, causing it to increase. Since the resistance to deformation in the *c* direction is less than that in the *a* direction, this effect predominates and the axial ratio increases to 1·856 as the number of electrons per atom increases to two for pure zinc.

Aluminium Alloys—Aluminium is a face-centred cubic trivalent metal and differs from the solutes already considered in that there is a large (symmetrical) overlap of electrons from the first Brillouin zone. Referring to *Figure 12*, this overlap occurs across the sloping faces of the zone. The first zone is almost full, with unoccupied states lying in thin layers parallel to the horizontal and vertical faces.[105] The aluminium lattice may therefore be regarded as symmetrically expanded by the presence of the symmetrical overlap, and we may expect solute elements of valency greater than three to give rise to a lattice expansion greater than would be given by size effects alone, while elements of valency lower than three would be expected to give somewhat lower expansions. The lattice spacings of the solid solutions of several elements in aluminium have been measured by AXON and HUME-ROTHERY,[106] and their interpretation of the results is of interest in emphasizing the complexity of the factors which must be considered. Part of this complexity arises from the fact that whereas the structures of the monovalent metals copper, silver, and gold may be regarded as analogous to an assembly of hard spheres (the ions) in contact, this is no longer true for aluminium, which has a larger amount of 'extra-ionic space'.

The lattice spacings of aluminium alloys, according to Axon and Hume-Rothery, may be interpreted in terms of four main factors, all of which may be simultaneously involved:

a The relative volume per valency electron in the solvent and solute crystals.

b The relative radii of the ions of solvent and solute metals.

c Brillouin zone considerations of the type discussed above.

d The electrochemical difference between solvent and solute metals.

Magnesium and germanium expand the aluminium lattice while lithium, silicon, copper, and zinc contract it, and for the purposes of

this discussion it is convenient to consider the apparent atomic diameters for these elements as deduced from the lattice spacing/composition curves by extrapolating to 100 per cent of solute. It will be appreciated that, were the spacing curves governed solely by atomic size relationships, this extrapolated value, or apparent atomic diameter, would check with the atomic diameter in the crystal of the solute element. It is found that the tetravalent elements silicon and germanium give apparent atomic diameters greater than the corresponding atomic diameters, while the monovalent elements lithium, copper, and silver give apparent diameters smaller than those characteristic of the pure crystal. This is in agreement with the principle that increasing the overlap should give rise to a relative expansion of the lattice compared with the distortion produced by atomic size effects alone, while the reverse should occur on decreasing the overlap. The regularity which would be expected on this simple theory is absent, however, and the other factors must be considered. The contraction is very marked for lithium, which we may consider as an example of the manner in which another factor operates. For lithium the ions are very small compared to the closest distance of approach of atoms in the body-centred cubic structure, and a large part of the volume of the metal is occupied by a relatively constant distribution of electrons. The volume per valency electron is large ($21{\cdot}5kX^3$), and the metal may be regarded as small Li^+ ions set in a very dilute valency electron gas. For aluminium, with three electrons per atom, the volume per valency electron is only $5{\cdot}51kX^3$, so that the electron gas is much more dense. It is also distributed less regularly, and the charge density is greater round the ions. This state of affairs may be qualitatively summed up by saying that the aluminium ions are strongly held together, and the atmosphere around an aluminium ion is relatively difficult to compress, while the atmosphere round an ion of lithium is comparatively easy to compress. When one of the compressible lithium atoms is substituted for the relatively less compressible aluminium atoms, the lattice forces are redistributed in such a manner that the electronic atmosphere round the lithium ion is compressed and the apparent atomic diameter is less than that in the crystal of lithium. Analysis of other results indicates that if the volume per electron for the solute is greater than that for the solvent the apparent atomic diameter tends to be smaller than the atomic diameter, while if the volume per electron for the solvent is

greater than that of the solute the apparent diameter of the solute is larger than in the pure crystal. This factor is complicated by the Brillouin zone effects and by the relative ionic radii of the components. If the ionic radius of the solute is much larger than that of the solute, the structure tends to be forced apart by the repulsive forces between ions with overlapping electron clouds, and the apparent atomic diameter will be greater than the crystal atomic diameter. In addition, when the solvent and the solute elements differ markedly in the electrochemical series, there is a tendency towards contraction of the lattice,[2] which will make the apparent atomic diameter smaller. In aluminium alloys, therefore, the effects of four separate factors, which may either reinforce or oppose each other, have been traced, so that lattice spacing relationships cannot be expected to be simple. The same principles will, of course, be operative in other solvents, and should be considered when examining the details of lattice expansion phenomena.

It will be clear from the discussion of lattice distortion in the monovalent metals, in magnesium, and in aluminium that the actual effect observed is the result of the interplay of many factors. It is therefore not surprising that Vegard's law is found to be a serious over-simplification.

Yet another effect may be encountered in lattice spacing curves, occurring in alloy systems containing a transition metal. The lattice spacings of the copper-nickel system, for instance, when plotted against composition give rise to a curve which consists of two distinct branches, intersecting at sixty atomic per cent of copper. According to the theories of Mott and Jones, nickel is characterized by a 3*d* band in which there is 0·6 vacancy per atom for electrons, and it is significant that the break in the spacing curve occurs at just that composition at which the valency electrons supplied by the monovalent copper are sufficient to fill up these vacancies. That such redistribution of electrons to fill up the vacant places for electrons in the outermost electron band of transition metals actually does occur in alloys is shown by the work of several authors on the x-ray emission and absorption spectra of alloys containing these metals.[107–9]

REFERENCES

[1] HUME-ROTHERY, W. *The Structure of Metals and Alloys* Inst. Metals Monograph and Report Series, No. 1, 1944
[2] —— *Atomic Theory for Students of Metallurgy* Inst. Metals Monograph and Report Series, No. 3, 1946
[3] RAYNOR, G. V. *An Introduction to the Electron Theory of Metals* Inst. Metals Monograph and Report Series, No. 4, 1947
[4] MOTT, N. F. and JONES, H. *The Theory of the Properties of Metals and Alloys* London, 1936
[5] HUME-ROTHERY, W., MABBOTT, G. W. and CHANNEL-EVANS, K. M. *Phil. Trans. roy. Soc. A* 233 (1934) 1
[6] GOLDSCHMIDT, V. M. *Z. phys. Chem.* 133 (1928) 397
[7] HUME-ROTHERY, W. and RAYNOR, G. V. *Phil. Mag.* 26 (1938) 129
[8] —— —— *Ibid* 26 (1938) 143
[9] PAULING, L. *J. Amer. chem. Soc.* 49 (1927) 765
[10] ZACHARIASEN, W. H. *Z. Krist.* 80 (1931) 137
[11] RAYNOR, G. V. *Phil. Mag.* 26 (1938) 152
[12] HUME-ROTHERY, W. *Ibid* 22 (1936) 1013
[13] OWEN, E. A. and ROBERTS, E. W. *Ibid* 27 (1939) 294
[14] —— and ROWLANDS, V. W. *J. Inst. Met.* 66 (1940) 361
[15] —— and ROBERTS, E. A. O'D. *Ibid* 71 (1945) 213
[16] —— *Ibid* 73 (1947) 471
[17] HUME-ROTHERY, W., LEWIN, G. F. and REYNOLDS, P. W. *Proc. roy. Soc. A* 157 (1936) 167
[18] COFFINBERRY, A. S. and HULTGREN, R. *Trans. Amer. Inst. min. met. Engrs* 128 (1938) 249
[19] RAYNOR, G. V. *J. Inst. Met.* 71 (1945) 553
[20] HUME-ROTHERY, W. and RAYNOR, G. V. *Ibid* 63 (1938) 201
[21] —— —— *Ibid* 63 (1938) 227
[22] RAYNOR, G. V. *Ibid* 66 (1940) 403
[23] HUME-ROTHERY, W. and RAYNOR, G. V. *Proc. roy. Soc. A* 174 (1940) 471
[24] KORNILOV, I. I. *Izvest. Akad. Nauk. S.S.S.R. Khim.* (1945) 104
[25] HUME-ROTHERY, W. and CHRISTIAN, J. W. *Phil. Mag.* 36 (1945) 835
[26] ZENER, C. *Amer. Inst. min. met. Engrs* Tech. Pub. No. 1856, 1946
[27] HUME-ROTHERY, W. *J. Inst. Met.* 35 (1926) 309
[28] WESTGREN, A. and PHRAGMÉN, G. *Metallwirtschaft* 7 (1928) 700
[29] WITTE, H. *Ibid* 16 (1937) 237
[30] EKMAN, W. *Z. phys. Chem. B* 12 (1931) 57
[31] HUME-ROTHERY, W., REYNOLDS, P. W. and RAYNOR, G. V. *J. Inst. Met.* 66 (1940) 191
[32] BRADLEY, A. J. and TAYLOR, A. *Proc. roy. Soc. A* 159 (1937) 56
[33] LIPSON, H. and TAYLOR, A. *Ibid* 173 (1939) 232
[34] BRADLEY, A. J. and GREGORY, C. H. *Phil. Mag.* 12 (1931) 143
[35] LAVES, F. and MOELLER, K. *Z. Metallk.* 29 (1937) 185
[36] JONES, H. *Proc. phys. Soc.* 49 (1937) 250
[37] —— *Proc. roy. Soc. A* 147 (1934) 400
[38] ZENER, C. *Phys. Rev.* 71 (1947) 846
[39] WEBB, W. *Ibid* 55 (1939) 297
[40] GOOD, W. A. *Ibid* 60 (1941) 605
[41] BARRETT, C. S. *Ibid* 72 (1947) 245; *cf* BARRETT, C. S. and TRAUTZ, O. R. *Metals Technology* 15 (1948) Tech. Pub. 2346
[42] HÄGG, G. *Z. phys. Chem. B* 12 (1931) 33
[43] ZINTL, E. and NEUMAYR, S. *Ibid* 20 (1933) 272

[44] ZINTL, E. and NEUMAYR, S. *Z. Elektrochem.* 39 (1933) 86
[45] WALLBAUM, H. J. *Z. Metallk.* 35 (1943) 218
[46] LAVES, F. and WITTE, H. *Metallwirtschaft* 14 (1935) 645
[47] DOERING, W. *Ibid* 14 (1935) 918
[48] LAVES, F. and WITTE, H. *Ibid* 15 (1936) 840
[49] —— —— *Ibid* 15 (1936) 15
[50] WITTE, H. *Z. Angew. Mineral* 1 (1937) 83
[51] —— *Ibid* 1 (1938) 255; *Metallwirtschaft* 18 (1939) 459
[52] SCHULTZE, G. E. R. *Z. Elektrochem.* 45 (1939) 849
[53] GORIA, C. and VENTURELLO, G. *Metallurgia Italiana* 32 (1940) 47
[54] LAVES, F., LOHBERG, K. and WITTE, H. *Metallwirtschaft* 14 (1935) 793
[55] STRAWBRIDGE, D. J., HUME-ROTHERY, W. and LITTLE, A. T. *J. Inst. Met.* 74 (1947) 191
[56] MAKAROV, E. S. *Izvest. Sekt. Fiziko-Khim. Anal.* 16 (1943) 149
[57] AGEEV, N. V. and MAKAROV, E. S. *Izvest. Akad. Nauk. S.S.S.R. Khim.* (1943) 87; *cf Compt. rend. (Doklady) Acad. Sci. U.R.S.S.* 38 (1943) 20
[58] —— —— *Izvest. Akad. Nauk. S.S.S.R. Khim.* (1943) 161
[59] MAKAROV, E. S. *Compt. rend. (Doklady) Acad. Sci. U.R.S.S.* 38 (1943) 191
[60] —— *Izvest. Akad. Nauk. S.S.S.R. Khim* (1943) 264; (1944) 29
[61] TENGNÉR, S. *Z. Anorg. Chem.* 239 (1938) 126
[62] MAKAROV, E. S. *Izvest. Akad. Nauk. S.S.S.R. Khim* (1946) 569
[63] CASTELLIZ, L. and HALLA, F. *Z. Metallk.* 35 (1943) 222
[64] MAKAROV, E. S. *Izvest. Akad. Nauk. S.S.S.R. Khim* 29 (1944) 114
[65] LAVES, F. and WALLBAUM, H. J. *Z. Angew. Mineral* 4 (1942) 17
[66] DEHLINGER, U. and NOWOTNY, H. *Z. Metallk.* 35 (1943) 151
[67] KLEMM, W. and WESTLINNING, H. *Z. Anorg. Chem.* 245 (1941) 365
[68] HUME-ROTHERY, W. and RAYNOR, G. V. *Phil. Mag.* 25 (1938) 335
[69] KUBASCHEWSKI, O. *Thermochemie der Legierungen* Berlin, 1943
[70] PAULING, L. *Phys. Rev.* 54 (1938) 899; see also PAULING, L. *J. Amer. chem. Soc.* 69 (1947) 542
[71] RAYNOR, G. V. *J. Inst. Met.* 70 (1944) 531
[72] BRADLEY, A. J. and TAYLOR, A. *Ibid* 66 (1940) 53
[73] RAYNOR, G. V. and PFEIL, P. C. L. *Ibid* 73 (1947) 397
[74] —— —— *Ibid* 73 (1947) 609
[75] —— and WALDRON, M. B. *Proc. roy. Soc. A* 194 (1948) 362
[76] —— and WAKEMAN, D. W. *Ibid* 190 (1947) 82
[77] —— —— *Phil. Mag.* 39 (1948) 245
[78] —— —— *J. Inst. Met.* 75 (1948) 131
[79] LEEMANN, W. G. *Aluminium-Archiv.* (1938) 9
[80] RAYNOR, G. V. and PFEIL, P. C. L. *Proc. roy. Soc. A* In press
[81] —— and PRATT, J. N. Unpublished work
[82] —— and LITTLE, K. *J. Inst. Met.* 71 (1945) 493
[83] LITTLE, A. T., RAYNOR, G. V. and HUME-ROTHERY, W. *J. Inst. Met.* 69 (1943) 423, 467
[84] RAYNOR, G. V. and HUME-ROTHERY, W. *Trans. Faraday Soc.* 44 (1948) 29
[85] BRADLEY, A. J. and GREGORY, C. H. *Proc. Manchester Phil. Soc.* 72 (1927–8) 91
[86] MOELLER, K. *Naturwissenschaft* 27 (1939) 167
[87] —— *Z. Metallk.* 31 (1939) 19
[88] —— *Ibid* 34 (1942) 171
[89] GEBHARDT, E. *Ibid* 32 (1940) 407
[90] ZAHAROVA, M. and SHTERNFELD, A. *Zh. tekh. Fiziki* (*J. tech. Physics*) 8 (1938) 2093
[91] RAYNOR, G. V. *Phil. Mag.* 39 (1948) 212

[93] GUERTLER, W. and ROSENTHAL, W. *Z. Metallk.* 24 (1932) 7
[93] RAYNOR, G. V. *Phil. Mag.* 39 (1948) 218
[94] VEGARD, L. *Z. Physik.* 5 (1921) 17
[95] WESTGREN, A. and ALMIN, A. *Z. phys. Chem. B* 5 (1929) 14
[96] HUME-ROTHERY, W., LEWIN, G. F. and REYNOLDS, P. W. *Proc. roy. Soc. A* 157 (1936) 167
[97] OWEN, E. A. and ROBERTS, E. W. *Phil. Mag.* 27 (1939) 294
[98] RAYNOR, G. V. Unpublished work
[99] BEARDEN, J. A. and FRIEDMANN, H. *Phys. Rev.* 58 (1940) 387
[100] FARINEAU, J. *J. phys. Radium* 10 (1939) 327
[101] MANNING, M. F. *Phys. Rev.* 55 (1939) 682
[102] JONES, H. *Proc. roy. Soc. A* 147 (1934) 400
[103] RAYNOR, G. V. *Ibid* 174 (1940) 457
[104] —— *Ibid* 180 (1942) 107
[105] MATYÁS, R. *Phil. Mag.* 39 (1948) 429: see also AXON, H. J. and HUME-ROTHERY, W. *Proc. roy. Soc. A* 193 (1948) 1
[106] AXON, H. J. and HUME-ROTHERY, W. *Ibid A* 193 (1948) 1
[107] FARINEAU, J. and MORAND, M. *J. phys. Radium* 9 (1938) 447
[108] BEARDEN, J. A. and BEAMAN, W. W. *Phys. Rev.* 58 (1940) 396
[109] FRIEDMAN, H. and BEAMAN, W. W. *Ibid* 58 (1940) 400
[110] DOUGLAS, A. M. B. *Nature, Lond.* 162 (1948) 566
[111] RAYNOR, G. V. and WALDRON, M. B. *Ibid* 162 (1948) 566
[112] —— —— *Phil. Mag.* In press
[113] —— and WAKEMAN, D. W. *Ibid* In press

2

THEORY OF DISLOCATIONS

A. H. Cottrell

THE PLASTIC deformation of metals can be discussed theoretically in two different ways. In the first,[1] a formal theory of plasticity is built up from empirical laws which idealize the observed plastic properties of metals; these laws are expressed mathematically and the attempt is made to solve the resulting equations under various conditions in order to analyse the behaviour of metals in complicated deformation processes, such as pressing, rolling *etc.* The second approach to the subject, which concerns us here, aims to construct a theory of the mechanics of the atomic movements involved in plastic deformation. While this molecular theory of plasticity is still at an early stage of development, it has already become clear that the essential link between the atomic structure of metals and their crystallographic plasticity properties (excluding certain cases of non-crystallographic plasticity, such as quasi-viscous flow along grain boundaries) is a particular kind of lattice defect known as a dislocation. The idea that plastic glide is due to the passage of dislocations through a crystal was first introduced in 1934 in simultaneous papers by TAYLOR,[2] OROWAN,[3] and POLANYI,[4] although the concept of a dislocation had been used earlier by PRANDTL[5] and DEHLINGER[6] in theories of mechanical hysteresis and crystal growth.

TYPES OF DISLOCATIONS

The concept of a dislocation arises naturally as a result of the crystallographic nature of plastic flow. The observation in strained metals of slip lines marking important crystal planes,[7,8] the proof that the observed plastic strains could be resolved into simple shears along these planes,[9] and the demonstration[10] that plastic flow is determined by the shear stress acting on the planes, showed clearly that the flow occurs by the sliding of certain atomic planes (glide or slip planes) across each other. Moreover, the structure in the glide planes remains crystalline during flow, since gliding occurs in the direction of closest atomic packing in the glide planes[8] and not in the direction of

maximum resolved shear stress. Accepting the picture given by these results, consider a plane A of atoms sliding in a certain crystallographic direction across a neighbouring plane B. Then, since atoms in a crystal are not rigidly bound to each other but are only elastically coupled, and since thermal vibrations and other sources of irregularity must make the forces acting over the glide planes non-uniform, we see that at any instant different portions of the plane A will, in general, have slipped over B by different amounts. Dislocations are defined as the boundaries between these different regions of the glide plane. It is to be noted that the existence of the discrete atomic structure in the glide

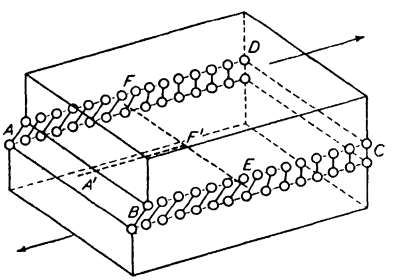


Figure 1. An edge dislocation

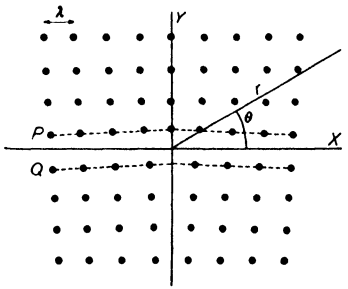


Figure 2. The structure of an edge dislocation

plane must limit the differences in slip amounts between neighbouring regions to the value of the atomic spacing, or possibly in some lattices to a small multiple or fraction of this spacing; we call the difference in slip amounts across a boundary the strength of the dislocation constituting that boundary. A second effect of the discrete structure is that all dislocations must be made up from a few standard types having well defined forms.

The first type is the Taylor or edge dislocation shown in *Figure 1*. In this the glide plane ABCD is divided by the dislocation line EF into the slipped region ABEF and the unslipped region FECD. The slip direction is A'F' and is perpendicular to the dislocation line. The atomic structure in an edge dislocation is shown in *Figure 2*, which is a section normal to EF, through *Figure 1*.

We see that the atoms in the upper half-crystal P are compressed along the slip direction and those in the lower half Q are extended. Dislocations can also exist which are the inverse of that shown in these diagrams *i.e.* dislocations in which the upper half-crystal is extended

and the lower half compressed. Accordingly, those of the type shown in *Figures 1* and *2* are called positive dislocations and their inverses are negative dislocations. Since a positive dislocation can always be turned into a negative one, and *vice versa*, by inverting the diagram or the crystal, this distinction would be trivial were it not for the important fact that the forces exerted between dislocations depend upon whether the dislocations are of the same sign or not.

The second fundamental type of dislocation is the screw dislocation introduced by BURGERS[11,12] and illustrated in *Figure 3*. In this, a part ABEF of the slip plane has slipped in the direction EF, while the

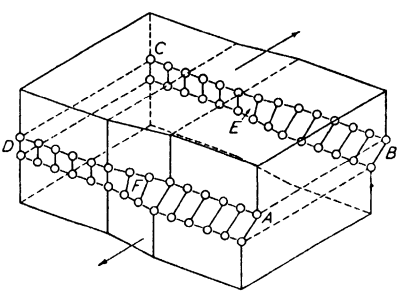


Figure 3. A screw dislocation

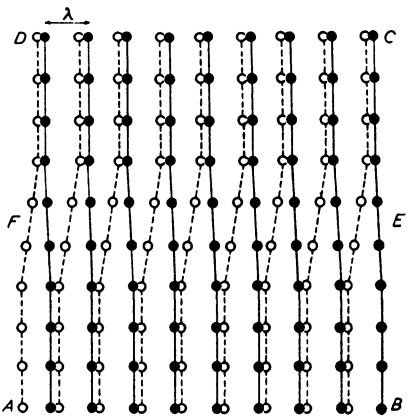


Figure 4. The structure of a screw dislocation

remainder FECD has not slipped, and the boundary EF constitutes the screw dislocation. It is to be noted that the line of the screw dislocation is in the slip direction and not perpendicular to it as in the previous case. In both cases further slip is produced by the migration of the dislocated region EF towards CD *i.e.* the slipped region ABEF grows at the expense of the unslipped region FECD. The dislocations differ, however, in the orientation of the direction of the dislocation motion relative to the slip direction. The atomic arrangement in a screw dislocation is shown in *Figure 4*, which is a plan view of part of the atomic planes which slip over each other; in this diagram the full lines and circles represent the atoms of the upper plane and the others are the lower plane. As with edge dislocations, screw dislocations of opposite signs, which are inverses of each other, can exist.

Although edge and screw dislocations are linear discontinuities,

dislocations of arbitrary non-linear form can easily be made by joining together segments of edge and screw dislocations, as in *Figure 5*, and it is probable that dislocations in actual crystals are of this compound type. BURGERS[11,12] has pointed out that a property of dislocation lines is that they can never end at an interior point in a crystal; they must either form closed chains of edge and screw segments or must end at the crystal surface. This is evident if we try to picture a dislocation which ends within a crystal. For example, suppose that the edge dislocation of *Figure 1* extends only from E to F'; by this we mean that the slipped region of the glide plane is only A'BEF', not ABEF. But this necessarily defines the line A'F' as a screw dislocation. Thus we cannot end an edge dislocation within the crystal without starting a screw dislocation, and *vice versa*.

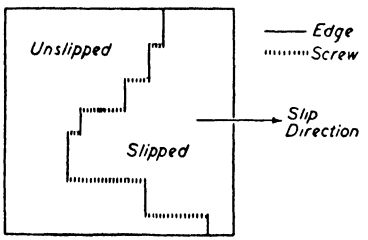


Figure 5. A compound dislocation formed from segments of edge and screw dislocations

Metals do not, of course, possess the simple cubic lattice we have been using, and it is more difficult to visualize the structures of dislocations in actual metal lattices. However, many important properties of dislocations depend upon the stress distribution about a dislocation at distances from its centre which are large compared with λ, and in an elastically isotropic medium the stress distribution at such distances depends only upon the strength of the dislocation and not on the detailed form of its centre. The strength of dislocation in a real crystal can be found by determining the distance in the slip direction between neighbouring equilibrium positions for the atoms of a slip plane.

Although dislocations are too small to be directly observable, as yet, in metals, so that our evidence for their existence must be indirect, it is interesting to note that in the two-dimensional 'crystals' of bubbles studied by BRAGG and NYE[13] plastic glide occurs by the passage of dislocations along the directions of closest packing. *Plate I* is an example of a dislocation in the bubble raft.

ELEMENTARY PROPERTIES OF DISLOCATIONS

The primary properties of a dislocation are its mobility along a glide plane and the fact that such movement causes the half-crystals on either

side of the glide plane to be displaced relative to each other in the slip direction. These properties are responsible for the importance of dislocations in plasticity and distinguish dislocations sharply from other types of lattice defects. Thus, for example, although it is a mobile defect, a vacant atomic site does not produce external change of shape of a crystal during its movements in the interior. It is quite easy to see qualitatively that a dislocation is highly mobile in the slip plane. In *Figure 6* the full circles represent the configuration of a dislocation initially centred about the line AA′. If the atoms in the upper row move forward in the indicated direction to the positions marked by the open circles, the dislocation has moved forward one interatomic spacing from AA′ to BB′. Each atom in the upper row is attracted by those of the lower row to the nearest site vertically above an atom of the lower row; as a result, when the atomic movements occur which move the dislocation from AA′ to BB′, atoms to the left of AA′ move under attractive forces from the lower plane and those to the right move against resisting forces. To a first approximation these forces cancel, so that the external force required to move the dislocation is vanishingly small. To a higher degree of approximation cancellation does not occur, and a small force is required to move the dislocation; the difficult problem of computing this force will be discussed later.

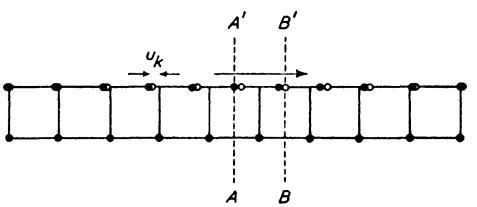


Figure 6. The movement of a dislocation

When a crystal is subjected to an external shear stress, acting on the glide planes in the glide direction, a force is produced on a dislocation tending to move it in that direction which enables the crystal to give way to the stress. This force is easily calculated by a method due to MOTT and NABARRO.[14] Consider a crystal (either *Figure 1* or *3*) in the form of a cube of edge length L and let it be subjected to a shear stress σ acting on the upper and lower faces in the direction of slip. Hence the external forces are σL^2. Then if the dislocation moves across the entire crystal, the crystal halves are displaced a distance λ relative to each other, so that the external forces do work equal to $\lambda\sigma L^2$. If the force per unit length on the dislocation is F, the total force is FL, and, in moving across the crystal with the dislocation, this force does work

equal to FL^2. Equating the work done gives the force on the dislocation as

$$F = \lambda\sigma \quad \ldots\ (1)$$

The displacement of the crystal halves and the shear produced by the movement of dislocations are easily found. Consider a crystal of width L_1 in the slip direction and thickness L_2 normal to the slip planes. If an edge type dislocation moves across the entire crystal it produces a displacement λ. Then if it moves a distance l, the corresponding displacement is $\lambda l/L_1$ and the average shear strain of the crystal is

$$\gamma = \lambda l/L_1L_2 \quad \ldots\ (2)$$

Similarly, if the velocity of the dislocation is $v = \mathrm{d}l/\mathrm{d}t$, the rate of shear of the crystal is

$$\mathrm{d}\gamma/\mathrm{d}t = \lambda v/L_1L_2 \quad \ldots\ (3)$$

In most problems we are concerned with crystals containing many dislocations, and it is convenient to define a quantity ρ, the density of dislocations, as the number of dislocation lines crossing perpendicularly a plane of unit area. Thus the number of dislocations in the present case is ρL_1L_2. If these each move an average distance l, the shear strain is

$$\gamma = \rho l\lambda \quad \ldots\ (4)$$

and if the average velocity is v, the rate of shear is

$$\mathrm{d}\gamma/\mathrm{d}t = \rho v\lambda \quad \ldots\ (5)$$

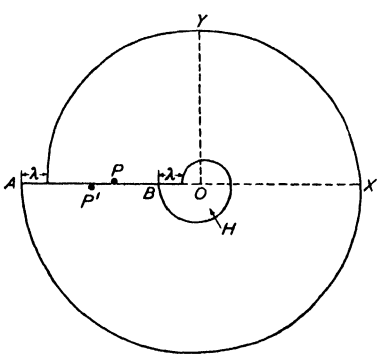


Figure 7. Elastic cylinder with a dislocation

STRESSES ROUND A DISLOCATION

Each dislocation is a centre of internal stress in a crystal. Thus in the edge dislocation of *Figure 2* the lower half-crystal exerts a force on the boundary P of the upper half-crystal, holding it in compression, and the upper half-crystal likewise holds the lower half in tension. TAYLOR[2] showed that these stresses could be determined by treating a dislocation as an example of a type of discontinuity which had already been analysed in classical elasticity theory by TIMPE, VOLTERRA, and others.[15,16,17] In this work an isotropic* elastic continuum of cylindrical form is considered, as shown in *Figure 7*. A dislocation is formed

* The effect of elastic anisotropy on the stresses round a dislocation is not considered here. This question has not yet been studied as fully as it deserves, but there are indications[11, 18] that anisotropy does not greatly alter the stress formulae.

by making a radial cut AO, sliding the cut surfaces along each other a distance λ, and joining them together again; thus points P and P', originally opposite each other on the plane AB, undergo a relative displacement λ in the radial direction. In a continuous medium the stresses would become infinite at the centre, so that it is necessary to imagine that a cylindrical hole H, with a radius of order of magnitude λ, is drilled out along the core of the dislocation. It is assumed that both the inner and outer surfaces of the medium are free from tractions and that the outer radius of the cylinder is infinite. The stresses set up by the dislocation can then be determined[11,19,20] for points not too near the centre, and by treating the problem as one in plane strain, *i.e.* we assume the dislocation is long in the z direction and suppose the z displacements to be zero, the normal stresses are:

$$\left.\begin{aligned} \sigma_{xx} &= -\frac{G\lambda}{2\pi(1-\nu)}\, y\, \frac{3x^2+y^2}{(x^2+y^2)^2} \\ \sigma_{yy} &= \frac{G\lambda}{2\pi(1-\nu)}\, y\, \frac{x^2-y^2}{(x^2+y^2)^2} \\ \sigma_{zz} &= \nu(\sigma_{xx}+\sigma_{yy}) \end{aligned}\right\} \quad \ldots\ldots \quad (6)$$

and the shear stresses are:

$$\left.\begin{aligned} \sigma_{xy} = \sigma_{yx} &= \frac{G\lambda}{2\pi(1-\nu)}\, x\, \frac{x^2-y^2}{(x^2+y^2)^2} \\ \sigma_{yz} = \sigma_{zy} &= \sigma_{xz} = \sigma_{zx} = 0 \end{aligned}\right\}$$

These formulae refer to either *Figure 2* or *Figure 7; G* is the shear modulus of the material, ν is Poisson's ratio, and λ is the dislocation strength. The analysis for a screw dislocation has been given by Burgers.[11] It will be observed from equations 6 that the stresses fall off inversely with distance from the dislocation, becoming zero at infinity and infinite at the origin. The region which must be excluded about the origin in which the stresses become too large for elasticity theory to apply is, however, quite small. Thus KOEHLER[19] excludes a cylindrical hole of radius r_0, defined by the condition that the maximum strain at the surface of the hole shall not exceed 0·1, and finds that r_0 is only about 6 Å for a dislocation of strength $\lambda = 2{\cdot}5$ Å. The infinity in the stresses at the origin which occurs when $r_0 = 0$ is due to the assumption that the medium is a continuum and does not, of course, exist for a dislocation in an atomic lattice. Stresses which are

continuous functions of the coordinates can only represent the actual state of affairs in a discrete structure when they vary slowly with position, relative to an atomic scale of size, in the crystal. Near the origin continuous stress functions lose their meaning, and one must discuss the displacements of individual atoms. The atomic displacements at the core of a dislocation are finite, and it is meaningless to talk of a variation of stress within the region bounded by these atoms. NABARRO,[18] following a method due to PEIERLS,[21] has been able to take account of the atomic structure in the glide plane and has developed a more complete theory which enables the structure of the core of the dislocation to be discussed. In this theory the equilibrium of the atoms in the plane P (*Figure 2*) is examined. These atoms are subjected to forces due to the atoms in the upper half-crystal which tend to spread the compression uniformly along the slip direction, and they are held back by forces from atoms in the lower half-crystal, particularly those from the plane Q. Equilibrium is given when these forces balance. The atomic structure of the medium is allowed for by using a periodic function of position, with period λ, to represent the forces from the atoms of the Q plane. The infinity in the stresses at the origin then disappears, for the factor $(x^2 + y^2)^2$ in the denominators of equations 6 is replaced by $\{x^2 + (y \pm \Upsilon)^2\}^2$, where the positive and negative signs are to be used in the upper ($y > 0$) and lower ($y < 0$) half-crystals respectively, and $\Upsilon = \nu\lambda/2(1 - \nu)$.*

The stresses due to a dislocation behave simply at infinity by falling off as $1/r$, where r is the distance from the origin, but several properties which involve an integration over the stress field diverge logarithmically. Thus consider the strain energy associated with the field of a dislocation. The strain energy in an element is determined by the product of stress and strain in that region, and since strain is proportional to stress, the energy must fall off as $1/r^2$. Integrating in polar coordinates over the entire field we have

$$\int_0^{2\pi} \int_{r_0}^{r_\infty} \frac{1}{r^2} r \, dr \, d\theta$$

which gives the total strain energy as a term of the form $\log (r_\infty / r_0)$, where r_∞ and r_0 are the radii of the outer and inner boundaries of the

* This is slightly different from the expression given in Nabarro's paper. I am indebted to B. A. Bilby for bringing the amendment to my notice.

medium. By an actual calculation of this type Koehler[19] obtains the 'self energy' W_s per unit length of a dislocation (excluding the energy of the highly strained region within r_0) as:

$$W_s = \frac{G\lambda^2}{4\pi(1-\nu)} \log \frac{r_\infty}{r_0} \qquad \ldots . \quad (7)$$

We see that a single dislocation in an infinite crystal should have an infinite strain energy. In a real crystal, on the other hand, it is reasonable to suppose that the action radius of a single dislocation is limited by the presence of other irregularities, and this is in fact what occurs for a pair of dislocations of opposite sign.

INTERACTION OF DISLOCATIONS

The strain energy of a pair of dislocations in a medium is not simply the sum of their self energies, even when we assume the stress at any point to be the sum of superposed stresses due to each dislocation acting by itself. Thus, if the stresses are σ_1 and σ_2, the strain energy is of the form:

$$(\sigma_1 + \sigma_2)^2 = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2$$

Integrating the first two terms gives the two self energies, but the cross-term $2\sigma_1\sigma_2$ gives a new energy depending upon the positions of the dislocations relative to each other. This term is an interaction energy between the dislocations and causes forces to be exerted between them. The energy and forces can be found[19] by integrating the strain energy over the entire field (excluding a cylindrical region of radius r_0 round each dislocation), but the integration of the cross-term is laborious. The interaction energy can be obtained more simply by a method which considers the work done against the forces due to an existing irregularity when a second irregularity is formed.[22] Consider in *Figure 8* a positive edge dislocation at the coordinate origin of a medium and suppose that a parallel dislocation, either positive or negative, is introduced at x_0, y_0. For purposes of calculation this second dislocation

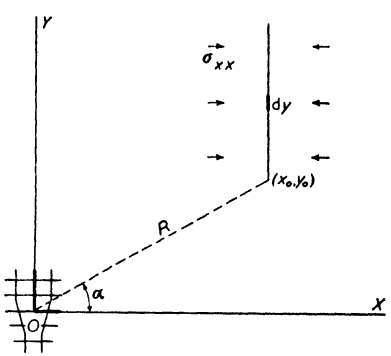


Figure 8. Work done when a second dislocation is introduced into the field of an existing dislocation

can be considered to be formed by making a cut parallel to the YZ plane extending from $y = y_0$ to $y = +\infty$, and by either inserting an extra half-plane of atoms in the cut (positive dislocation) or by removing a half-plane (negative dislocation). In such a process the boundaries of the cut region are displaced in the direction of the X axis and work is done by the X component of force due to the stress field of the first dislocation. Work is also done against the normal elastic resistance of the medium, but this will constitute the self energy of the second dislocation which does not interest us here. Clearly the interaction energy results from the work done by the forces from the first dislocation, and we may proceed to calculate this. For unit thickness in the Z direction, the work done by the forces acting on an element of area dy when a negative dislocation is formed is approximately $-\sigma_{xx}\lambda dy$ and the change in energy of the system due to this is $\sigma_{xx}\lambda dy$. Hence the interaction energy V is, from equations 6:

$$\int_{y_0}^{y_\infty} \lambda\sigma_{xx}dy = -\frac{G\lambda^2}{2\pi(1-\nu)}\int_{y_0}^{y_\infty}\frac{3x_0{}^2y + y^3}{(x_0{}^2+y^2)^2}\,dy$$

$$= -\frac{G\lambda^2}{2\pi(1-\nu)}\left[\log\,(x_0{}^2+y^2)^{1/2} - \frac{x_0{}^2}{x_0{}^2+y^2}\right]_{y_0}^{y_\infty}$$

and if we put $R^2 = x_0{}^2 + y_0{}^2$, $\cos\alpha = x_0/R$, let $y_\infty \to \infty$ and let $(x_0{}^2 + y_\infty{}^2) \to r_\infty{}^2$, this becomes

$$V = -\frac{G\lambda^2}{2\pi(1-\nu)}\,(\log r_\infty - \log R + \cos^2\alpha) \quad . \quad . \quad . \quad . \quad (8)$$

We thus obtain the radial force F_R and the tangential force F_α as:

$$\left.\begin{aligned} F_R &= -\frac{\partial V}{\partial R} = -\frac{G\lambda^2}{2\pi(1-\nu)}\frac{1}{R} \\ F_\alpha &= -\frac{\partial V}{R\partial\alpha} = -\frac{G\lambda^2}{2\pi(1-\nu)}\frac{\sin 2\alpha}{R} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (9)$$

Hence dislocations of opposite sign attract each other with a force that varies as the reciprocal of the distance between them. For dislocations of the same sign, the above expressions are reversed in sign and the dislocations repel. These conclusions were first arrived at by Taylor.[2] If we add the self energies (equation 7) to the interaction energy we obtain the total strain energy of a pair of dislocations near

the centre of a large cylindrical crystal of radius r_∞. For dislocations of opposite sign this gives:

$$2W_s + V = \frac{G\lambda^2}{2\pi(1-\nu)} \left(\log \frac{R}{r_0} - \cos^2 \alpha\right) \quad . \quad . \quad . \quad . \quad (10)$$

The log r_∞ term is no longer present in the energy, so that pairs of dislocations of opposite sign have finite strain energies. The interpretation of this is that, at very large distances from a dislocation pair, the superposed stresses σ_1 and σ_2 cancel and leave the material stress free; the strain energy is thus localized to the neighbourhood of the dislocations.

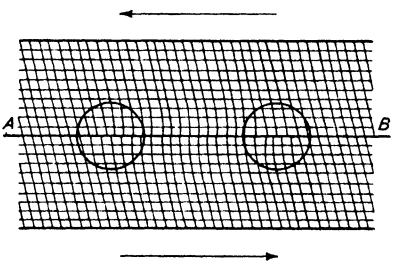


Figure 9. Dislocations of opposite sign in the same slip plane (After Orowan)

The special case of dislocations of opposite sign on the same slip plane (*Figure 9*) is important for considering the coalescence and mutual annihilation of dislocations of opposite sign and also the possibility of forming dislocations, in pairs, on a slip plane by local gliding[3] in a crystal under shear stress. For such dislocations there is no tangential force, and the force between them along the slip plane is F_R of equation 9.

Table I. Force between Dislocations of Opposite Sign on the same Slip Plane

Distance R *between dislocations (cm)*	10^{-7}	10^{-5}	10^{-3}
Force F *per unit length between dislocations (dyne cm^{-1})*	600	6	0·06
Equivalent shear stress $\sigma = F/\lambda$ *(gm mm^{-2})*	$2{\cdot}5 \times 10^5$	$2{\cdot}5 \times 10^3$	25

Taking $G = 4 \times 10^{11}$ dyne cm^{-2}, $\nu = 0{\cdot}34$, and $\lambda = 2{\cdot}5 \times 10^{-8}$ cm, which are values typical of copper, the results in *Table I* are obtained. The equivalent shear stress given here is that stress applied to the crystal which would exert a force on a dislocation equal to that produced by the other dislocation, and is calculated from equation 1. We see that in the absence of other stresses in the crystal the dislocations will run together under the attractive force and annihilate themselves provided that they are close enough to overcome the constraining force

anchoring them to equilibrium positions between lattice rows, an aspect discussed in the next section. This attractive force can be balanced by an external shear stress acting in that direction which drives the dislocations apart; this stress is equal and opposite to the equivalent shear stress given by *Table I.* It is important to notice that the equilibrium is an unstable one. If, for example, the applied stress is $2{\cdot}5 \times 10^3$ gm mm^{-2} the dislocations will run together if they are closer than 10^{-5} cm and will separate if they are farther apart than this distance.

For dislocations to be generated in pairs in a crystal under an external shear stress σ, they must acquire an activation energy sufficient to separate them to at least the critical distance corresponding to unstable equilibrium, otherwise the external stress will be unable to hold them apart. Since the critical distance depends on the value of the applied stress, it follows that the activation energy also depends on this quantity. The activation energy is not given simply by equation 10, for the separation of the dislocations to a distance R causes a displacement of the half-crystals and work is done by the external forces. This work must be subtracted from equation 10. If we consider a crystal of unit thickness in the Z direction and width L along the slip direction, the formation of a dislocation pair and their separation by a distance R causes a displacement $R\lambda/L$ and work is done equal to $\sigma L . R\lambda/L = \sigma R\lambda$. Since for equilibrium $\sigma = -F_R/\lambda$, the work done by the external forces is $G\lambda^2/2\pi(1-\nu)$, and hence, subtracting from equation 10, the activation energy per unit thickness is:

$$W = \frac{G\lambda^2}{2\pi(1-\nu)}\left(\log\frac{R}{r_0} - 2\right) \qquad \ldots\ (11)$$

Using the previous numerical values and taking $r_0 = 6 \times 10^{-8}$ cm, this gives the results shown in *Table II.*

Table II. Activation Energy for the Formation of Dislocation Pairs by Local Gliding

Applied stress σ (*gm mm*$^{-2}$)	25,000	2,500	250	25
Critical spacing R (*cm*) ..	10^{-6}	10^{-5}	10^{-4}	10^{-3}
Activation energy (*electron volts per atom plane*)	0·77	2·9	5·1	7·3

In calculating these energies we have so far excluded from consideration the energy of the highly strained regions within the dislocation cores. Although a rigorous calculation of this energy is not possible a rough estimate can be obtained in several ways, and in each case a value of the order of one electron volt per dislocation per atom plane is obtained. The strain at a distance r from a dislocation is of order $\lambda/2\pi r(1-\nu)$. KOEHLER[19] excludes a region $r_0 = 6 \times 10^{-8}$ cm about a dislocation of strength $\lambda = 2{\cdot}5 \times 10^{-8}$ cm; taking the strain at half this distance to be the average throughout the excluded region, the mean strain is about 0·2. Assuming Hooke's law to apply in this region, the mean stress is about $0{\cdot}2G$ and the strain energy is:

$$\tfrac{1}{2} \times \text{stress} \times \text{strain} \times \text{volume} = 0{\cdot}02G\pi r_0^2 d$$

where d is the interatomic spacing along the dislocation. Taking $G = 4 \times 10^{11}$, $r_0 = 6 \times 10^{-8}$, $d = 2{\cdot}5 \times 10^{-8}$, the core energy per dislocation per atom plane is about 1·4 electron volts. Other estimates have been given by BRAGG[23] and HUNTINGTON.[24] Bragg, using the criterion that the strain energy within the excluded region cannot exceed the latent heat of melting, obtains an energy of about half an electron volt, whereas Huntington, evaluating the energy due to the short range electrostatic forces between the atoms at the core of a dislocation in a rock-salt crystal, obtains a value of about one electron volt.

We see then that the activation energies given in *Table II* for the formation of a pair of dislocations have to be increased by about two electron volts. For annealed single crystals of pure metals the shear stress for yielding is about 10 to 50 gm mm^{-2}; hardening can raise this value a thousandfold. Thus for all cases in practice the activation energy is a few electron volts for each atomic plane normal to the dislocation. The formation of a pair of dislocations, even quite short ones ($\simeq$ 100 atoms long), would therefore require a few hundred electron volts. Activation energies as large as this cannot be produced by thermal fluctuations, and so we must discount the possibility of the production of dislocation pairs by thermal fluctuations in a crystal under stress.

A more detailed theory of dislocation pairs has been given by Nabarro.[18] Koehler[19] has also considered the possibility of the formation of dislocations at free surfaces. Again the activation energy is

large (about one half the above values), and there is thus no more possibility of their formation by thermal fluctuations than in the above example.

THE SHEAR STRESS REQUIRED TO MOVE A DISLOCATION

In the theory developed above, the energy of a single dislocation (remote from free surfaces, dislocations, or other irregularities) is independent of its position, which means that the dislocation ought to be able to move under the smallest external shear stress. This is in accord with the qualitative discussion of the mobility of dislocations, where it was shown that, to a first approximation, the interatomic forces assisting or resisting the external force balanced each other. More exactly, however, the effect of the atomic structure must be to make the energy of a dislocation depend upon its exact position *i.e.* whether the symmetry plane (AA′ of *Figure 6*) marking the dislocation centre passes through a plane of atoms or not. This energy must vary periodically as the dislocation moves, so that at least one position of stable equilibrium occurs in every interval of one atomic spacing along the line of motion of the dislocation. Hence a force will constrain the dislocation to an equilibrium position and the external force must exceed this constraining force before slip can occur. That it will be very small is clear, because first approximation treatments give it as zero; for the same reason, because it is necessary to proceed to a higher degree of approximation before the existence of the force can be demonstrated, even a rough calculation of the magnitude involves a delicate analysis. The problem was first attempted by PEIERLS,[21] and his calculations were later extended by NABARRO.[18]

This work has shown that an important property of a dislocation affecting the restraining force is its width, that is, for an edge dislocation, the extent along the slip direction of the heavily strained region. The width is controlled by the ratio of the forces acting across the slip plane to the forces acting in each half crystal. Thus, in *Figure 2*, the atoms in the plane P are under two forces:

1. The compressed atoms in the upper half-crystal try to spread the compression uniformly along the plane P and widen the dislocation, and
2. the atoms in the lower half-crystal, and particularly those in the adjacent layer Q, try to pull the atoms in P into alignment and make the dislocation narrow.

The equilibrium width is that at which the forces balance. According to Peierls and Nabarro, the alignment forces are very strong and so the dislocation is extremely narrow, the region in which the displacements are about one quarter of the lattice spacing being itself not appreciably greater than one lattice spacing; however, it is admitted by the authors that their calculations may underestimate the width.

The constraining force is very sensitive to the width of the dislocation and falls rapidly as the dislocation becomes wider. This can be seen in a general way if we consider a limiting case in which the distortion is spread uniformly along the entire slip direction *i.e.* a case in which a row of N atoms of uniform spacing $L/(N-1)$ is matched against a row of $N+1$ atoms of uniform spacing L/N. Then, if N is very large, the interaction energy of the lines will hardly alter as the one row is moved across the other; in the limit $N = \infty$, the interaction energy will be constant and the restraining force will be zero. According to Nabarro's calculation, the stress required to move the dislocation by overcoming the restraining force is:

$$\sigma_t \frac{4\pi}{1-\nu} e^{-2\pi/(1-\nu)} \quad \ldots . \quad (12)$$

where σ_t is the theoretical shear stress for a perfect lattice ($\simeq G/10$) and ν and G are as previously defined. This gives a critical shear stress of about 1,000 gm mm^2, which is ten to fifty times larger than the observed values for annealed pure single crystals. However, it is pointed out that the approximations used in the calculation probably cause the width of the dislocation to be underestimated by a factor of two or three, in which case the critical shear stress is overestimated by a factor of 1,000 or more. The difficulties of making an estimate are so great that the only conclusion one can make is that the constraining stress is probably less than the observed elastic limits of metal crystals.

The existence of a constraining force slightly modifies the conclusions arrived at earlier (page 87) concerning the mutual attraction of dislocations of opposite sign on the same slip plane. Evidently, if the dislocations are separated by a distance greater than that at which the constraining force equals the interaction force, the dislocations will not run together. Nabarro estimates this distance as about 10,000 atomic spacings, although it is admitted that this result, which depends on the value of the constraining force, is no more accurate than the

latter. It is also suggested that, as the spacing of 10,000 is about the same as the width of mosaic blocks in crystals, its existence may be the reason for the appearance of this characteristic length.

ASSEMBLIES OF DISLOCATIONS

In a certain sense, a dislocation is a very economical lattice defect; the introduction of a dislocation is the simplest way, involving the least expenditure of energy, of arranging a misalignment between crystal halves. It is therefore reasonable to suppose that the more

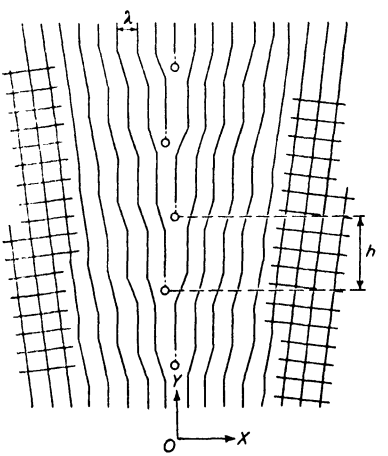


Figure 10. A transition surface between crystallites formed by a set of parallel edge dislocations

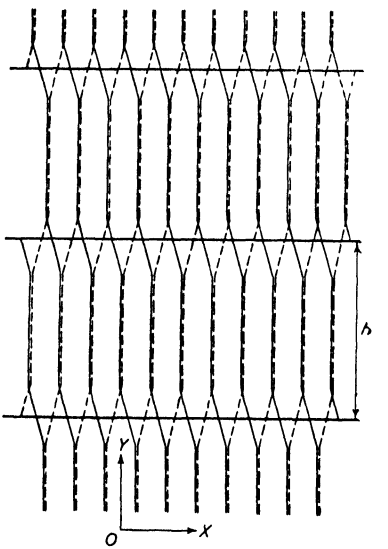


Figure 11. A transition surface composed of screw dislocations

macroscopic irregularities in crystals, *e.g.* mosaic structure and crystal fragmentation in the cold worked state, are built up from assemblies of dislocations. The simplest assemblies to consider are those which could form transition surfaces between mosaic blocks or crystallites inclined at small angles to each other; these have been considered in detail by BURGERS[11,12] and discussed also by TAYLOR,[25] BRAGG,[26] and LENNARD-JONES.[27] *Figure 10* shows the very simplest case of a transition surface in which two crystallites are joined by a set of parallel edge dislocations of the same sign, regularly spaced at intervals h over the plane OYZ. The crystallites are rotated with respect to each other through an angle $\tan^{-1}\lambda/h$ about an axis parallel to the z axis. Burgers has shown that in

such a case the crystallites are free of stress at large distances from the transition surface compared with h. A rotation about the y axis is obtained when the dislocations are arranged parallel to this axis, and, by considering two intersecting sets of edge dislocations in the transition surfaces, parallel to the y and z axes respectively, relative rotation of the crystallites about any arbitrary axis confined to the yz plane is obtained. Such rotations are not the most general, however, as they do not include a component about the x axis. No arrangement of edge dislocations in the transition surface can produce this, but it is obtained when screw dislocations are introduced. *Figure 11* shows a simple transition surface, composed of screw dislocations, between two crystals which are oppositely sheared; by introducing a second set of screw dislocations intersecting the first set, a rotation may be obtained.

Thus it is geometrically possible to join two crystals in arbitrary orientation to each other by means of a complex transition surface composed of sets of edge and screw dislocations. The stability of such surfaces has not so far been fully studied, although Burgers has considered the simpler cases. In connection with *Figure 10* one might be tempted to argue qualitatively that dislocations of the same sign repel each other and conclude that the transition surface should break up spontaneously, the dislocations driving each other away along the glide planes of the adjoining crystallites. This conclusion is, however, wrong. In deriving the stress formulae of equations 6 and the subsequent interaction formulae, the picture assumed is of one or a few dislocations in an otherwise perfect single crystal of infinite extent; on the other hand the structure in *Figure 10* represents part of two crystals joined at an angle to each other and of either finite or infinite extent. The assumptions of the previous theory restrict the discussion to the behaviour of a dislocation introduced into a lattice in which the atomic slip 'planes' are planes in the strict geometric sense. In finite crystals or polycrystals, dislocations in 'bent' atomic slip planes have to be considered. That the condition of finiteness must modify the behaviour is evident from the fact that the stresses given by equations 6 only become zero at infinity, which is incompatible with the boundary conditions for a finite crystal free from external forces. The argument can be made more specific by considering qualitatively the effect of a positive edge dislocation in a crystal of finite thickness h in the y direction as in *Figure 12a*. The dislocation, centred at the origin, divides the

crystal into halves on either side of the slip plane $y = 0$; the upper half tries to extend itself in the slip direction but is restrained by the lower half, which tries to contract. This is analogous to a bimetallic strip, and in the absence of external forces the crystal will become bent, as in *Figure 12b*. An equilibrium arrangement will result when the tendency of the material near the dislocation (between A and A′) to bend is

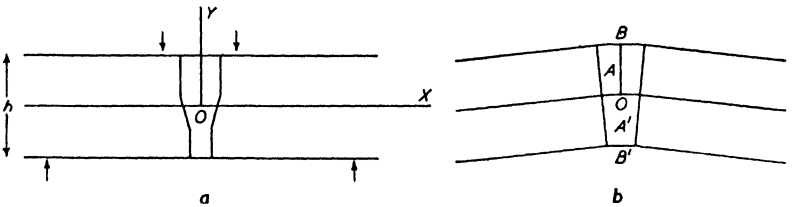


Figure 12. Bending of a finite crystal by a dislocation

balanced by the tendency of the more remote regions (from A to B and A′ to B′) to remain undeformed. The angle of bend is decreased as h is increased, and in the limit $h = \infty$ it becomes zero and the case discussed earlier (page 83) is obtained. It can be seen qualitatively that σ_{xx} is a tension stress in the region A to B and a compression from A′ to B′; hence, if a positive dislocation is introduced into AB by inserting an extra part plane of atoms into the yz plane, or into A′ B′ by removing a part plane, the stress does work and the interaction energy between the dislocations is negative. Thus, in a finite crystal free from external restraints, dislocations of the same sign will tend to assemble in transition surfaces.

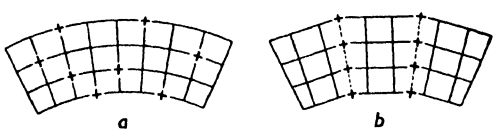


Figure 13. a Bending of a crystal by randomly distributed dislocations of the same sign. b Assembly of the dislocations into sheets, producing crystallites

Experimental evidence supporting these suggestions has been obtained by CAHN.[28] A characteristic of some crystals which have been bent plastically is that, on subsequent annealing, their x-ray Laue asterisms split into discrete spots. The interpretation of this is that the bending produces bent glide planes and that annealing allows the crystal to eliminate this bending by forming crystallites (polygonization). The dislocation theory picture of this process is illustrated by *Figure 13*. In the first diagram, bending of the lattice is produced by randomly distributed dislocations of the same sign, represented by the crosses. On annealing, these dislocations assemble in sheets, producing

crystallites inclined at small angles to each other, as in the second diagram. If this is true, one might expect to obtain microscopical evidence for these transition surfaces and they ought to be arranged roughly perpendicularly to the plane and direction of glide. Such evidence has been obtained by Cahn and is shown in *Plate III.* These photographs are of zinc crystals which, after bending and annealing, were sectioned, etched, and examined microscopically. Boundary lines are developed by this treatment which are orientated, relative to the glide planes and the axis of bending, in the positions expected for transition surfaces.

Evidence for transition surfaces composed of dislocations has also been obtained by BRAGG and NYE,[13] who have shown that in two-dimen-mensional crystals of bubbles the equivalents of such surfaces often appear. *Plate II* is an example.

SEGREGATION OF SOLUTE ATOMS ROUND DISLOCATIONS

Due to its stress field, a dislocation will interact with other irregularities which are sources of internal stress in a crystal, such as particles of precipitate, crystallite boundaries, and other dislocations. In general, the more advanced problems of the theory of dislocations are concerned with this interaction, and it is by their solution that the processes of hardening and softening are to be understood. However, there is one case, the segregation of solute atoms round dislocations, which is rather more straightforward and may usefully be considered first.

As *Figure 2* shows, atoms above a positive edge dislocation are compressed, whereas those below are expanded. The strain energy due to this distortion can be relieved if the natural lattice parameter of the upper region can be reduced locally and that of the lower region increased; the strains will then remain but not the stresses. It has been shown recently[22] that such a state can be obtained by a segregation of solute atoms round a dislocation. In substitutional solutions large solute atoms will collect in the expanded region and small ones in the compressed region. If an atom in interstitial solution causes a local lattice expansion it will migrate to the expanded region, and *vice versa.* However, where interstitial atoms produce non-symmetrical local distortions, the segregation process may cause shear stresses to be relieved,[29] in addition to the hydrostatic stresses relieved by the dilatation change. In most cases, atoms in substitutional solution produce

symmetrical distortions and so can only relieve hydrostatic stresses. It has been shown [22] for such a case that a solute atom of radius $r_a(1 + \varepsilon)$ in a solvent of atomic radius r_a which is under stress has an interaction energy with the stress field:

$$V = -\frac{4}{3}\pi\varepsilon r_a^3 (\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) \quad \ldots . \quad (13)$$

where σ_{xx}, σ_{yy} and σ_{zz} are the normal stresses of the field. Substituting from equations 6 and using the polar coordinates of *Figure 2*, the interaction energy with a positive edge dislocation is

$$V = \frac{4}{3}\frac{1+\nu}{1-\nu} G\varepsilon r_a^3 \lambda \frac{\sin\theta}{r} \quad \ldots . \quad (14)$$

where G, λ, and ν retain their usual meaning. We notice that V is positive on the upper side ($0 < \theta < \pi$) of the dislocation for a large solute atom ($\varepsilon > 0$) and negative on the other side, which agrees with the qualitative picture of a large atom being repelled from the compressed region and attracted into the expanded one. Since V is a function of r and θ, both radial ($-\partial V/\partial r$) and tangential ($-\partial V/r\partial\theta$) forces are exerted between the atom and the dislocation. The magnitude of V is easily examined. For a solute which enters easily into solution a value for $\varepsilon = 0{\cdot}05$ is reasonable. Using typical values for the constants and taking $\theta = \pi/2$, V is equal to kT at about 6×10^{-8} cm from the dislocation centre. This shows that the effect can produce a marked local segregation of solute atoms and thus alter the properties of the dislocation.

If time for atomic migration is allowed, an equilibrium distribution or atmosphere of solute atoms must build up round a stationary dislocation to a density determined by V. For dilute atmospheres this density is

$$c = c_0 \, e^{-V/kT} \quad \ldots . \quad (15)$$

where c_0 is the average concentration. This formula cannot apply to dense atmospheres, however; an atmosphere can become saturated when sufficient solute atoms to relieve the hydrostatic stresses completely are contained in it, for then there is no tendency for further solute atoms to migrate to the dislocation. It is important to notice that extremely small amounts of solute are required for atmospheres. For a metal in which the dislocation density is as high as possible (10^{12} cm^{-2}, see page 104) only about 0·1 per cent of solute is required to provide one solute atom per dislocation per atom plane. In annealed

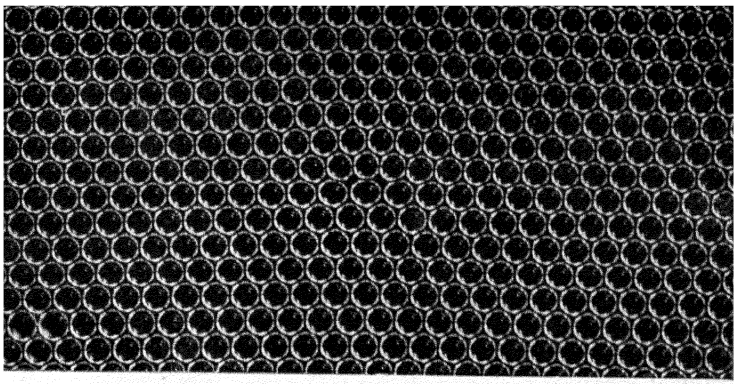


Plate I. A dislocation in a two-dimensional bubble raft

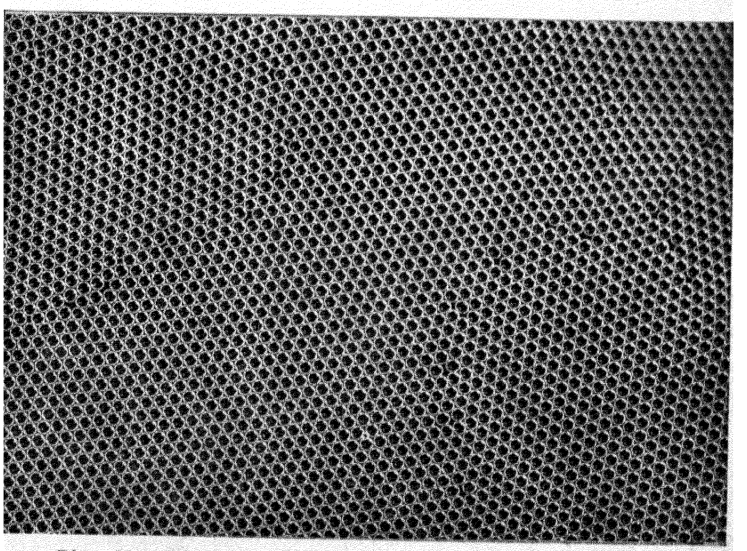


Plate II. A transition surface between crystals in the bubble raft

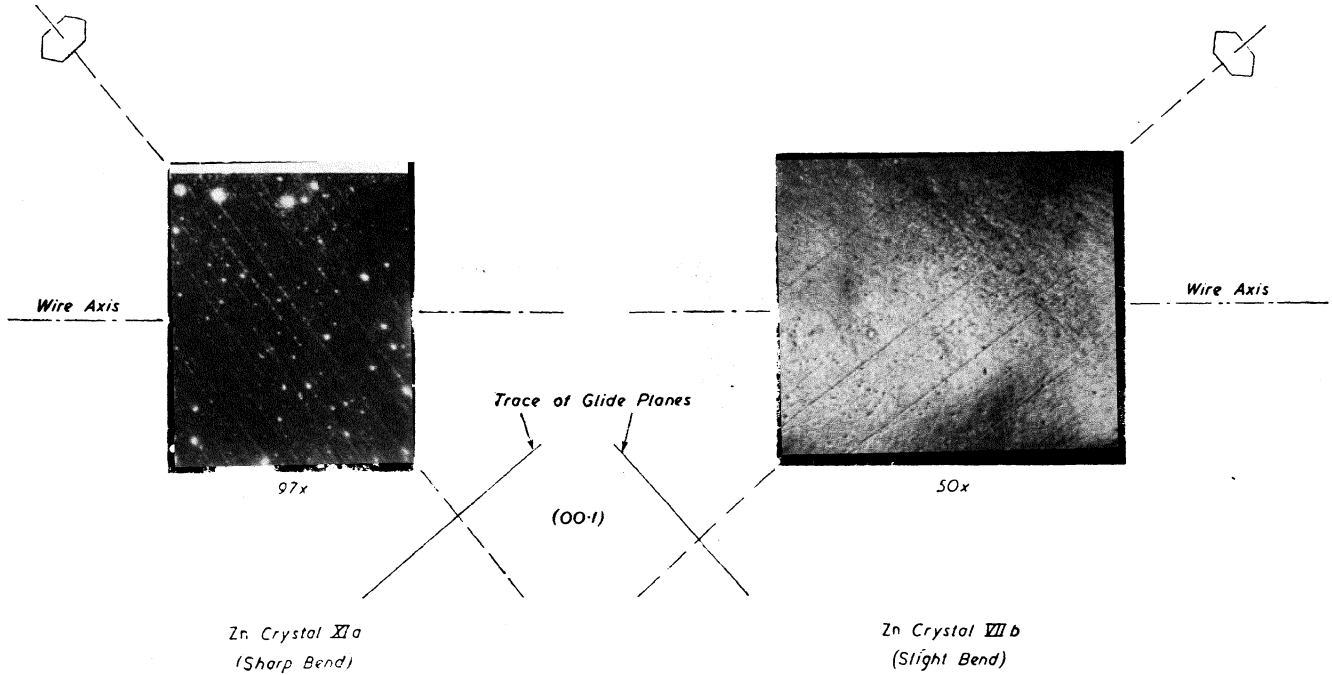


Plate III. Boundaries developed in polished and etched zinc crystals in the positions expected for transition surfaces. The plane of bending and the hexad axis are in the plane of the figure

metals the dislocation density must be far lower, and in such cases there can be no shortage of solute atoms for dislocations even in so-called pure metals (99·999 per cent).

This theory can be used to explain the well known effect of plastic deformation in producing rapid and localized precipitation on the slip planes in supersaturated solid solutions. The force attracting solute atoms to a dislocation causes a systematic migration of the atoms to a common centre, and this must be a more rapid process than the chance aggregation of randomly migrating atoms, so that precipitation should occur first in dislocations.

The other effects to be expected are concerned with the attempt to produce plastic flow by moving dislocations which are surrounded by atmospheres. Under an external force a dislocation will start to move and so to leave its atmosphere behind; the energy of the atmosphere is then raised, which means that a force is exerted anchoring the dislocation to the atmosphere. If the applied force is insufficient to overcome the anchoring force, the dislocation cannot escape. It can still move, however, by a process in which the solute atoms migrate with it, and this will be controlled by the diffusion rate of the solute. It can be shown that this slow dislocation process is fast enough to operate in slow creep, and the suggestions have been made that it occurs in micro-creep[22] and in the blue-brittleness of iron.[29] Micro-creep, discovered by CHALMERS[30] when studying tin crystals, is a slow creep occurring at very low stresses. It has the characteristics that the initial strain rate is proportional to the stress, being zero at zero stress, and that the strain rate decreases with increasing strain, becoming almost zero after a total strain of about 10^{-5}. These effects are consistent with the above theory. It can be shown[31] that, for small stresses, a linear strain rate/stress relation should be observed. Also, if the creep is due to dislocations initially present in the specimen and these dislocations move too slowly to create others (page 100), they will gradually be used up, and the strain rate should decay with strain in the observed manner.

To produce rapid flow the dislocations must be freed from their atmospheres by the application of a force which exceeds the atmosphere binding force. The important feature here is that the dislocations, once freed, will be under unnecessarily large external forces and should accelerate markedly when rapid flow under smaller forces becomes possible. This resembles the phenomenon of the sharp yield point,

which is a prominent feature of the mechanical behaviour of iron and soft steel, and is also observed to a lesser degree in certain other materials.[32] Several investigations[33,34,35] have shown that the yield point of iron is associated with the presence of small amounts of carbon in the metal, which suggests that in this case the atmospheres are composed of carbon. This is consistent with the theory, because the carbon atom greatly distorts the α iron lattice and this must cause a large interaction energy and yield point. A recent estimate[36] of the force required to free a dislocation from a carbon atmosphere in α iron gives a yield point of the right order of magnitude. Strain ageing in iron is also accounted for by the theory. A freshly strained iron specimen shows no yield point, but if the iron is rested for a certain period of time (dependent upon the temperature) the yield point returns. This is interpreted in terms of the migration of carbon atoms to the freed dislocations to form new atmospheres. NABARRO[29] has shown that the activation energy for strain ageing in iron is the same as for the diffusion of carbon and that the characteristic time for the process is of the same order as would be expected for atmosphere formation.

FAST DISLOCATIONS AND SLIP BAND FORMATION

As mentioned above, a free dislocation under an unnecessarily large force must accelerate to a high velocity, the surplus energy from the work done by the force being stored as kinetic energy of motion. There can be little doubt that such dislocations are responsible for rapid slip in which individual slip bands are formed in very small time intervals. It must not be assumed, however, that a fast dislocation will accelerate indefinitely, for two effects prevent it from attaining the velocity of sound in the material. In the first place, as the atoms in the dislocation core slide past each other they move in the periodic force field exerted between the planes P and Q (*Figure 2*), and this must cause an oscillatory motion in the y direction in addition to the translation along the x axis. The bumping of atoms facing each other across the slip plane, caused by this motion, dissipates energy in the form of elastic waves transmitted from the dislocation to the rest of the crystal, and the amount of energy dissipated at each step increases with the dislocation velocity. It may therefore rise to such a value that all the surplus energy derived from the external force is dissipated, in which case the

dislocation will no longer accelerate. The process has been briefly discussed by OROWAN[37] and FRENKEL and KONTOROVA.[38]

The second effect, deduced independently by FRENKEL and KONTOROVA[38] and FRANK,[39] is in some ways reminiscent of the relativistic behaviour of fast particles. The complexity of the dynamics of fast dislocations makes it necessary to use a simplified model of a dislocation for discussing the effect. Suppose, therefore, that in *Figure 6* an upper row of flexibly coupled atoms is sliding over a lower row of fixed and uniformly spaced atoms, and let u_k be the displacement in the slip direction of the kth atom of the upper row. The potential energy of this atom is determined by its distance from its neighbours in the upper row and by its position in the periodic field produced by the lower row. Assuming elastic forces obeying Hooke's law between the atoms of the upper row and a sinusoidal field from the lower row, the potential energy of the system can be written in the form:

$$U = \sum_k A\left(1 - \cos\frac{2\pi u_k}{d}\right) + \tfrac{1}{2}\alpha\sum_k\left(u_{k+1} - u_k\right)^2 \quad \ldots . (16)$$

where d is the lattice spacing along the row, α is the coefficient of the elastic force between atoms of the upper row, and A is the amplitude of the periodic field produced by the lower row. The equation of motion of the kth atom is then:

$$m\frac{\mathrm{d}^2u_k}{\mathrm{d}t^2} = -\frac{\partial U}{\partial u_k} \quad \ldots . (17)$$

where m is the atomic mass. This equation can be solved, subject to simplifying conditions, giving:

$$u_k = \frac{2d}{\pi}\tan^{-1}\left\{C_k\exp\left(\pm\tfrac{2\pi}{d}\sqrt{\left(\tfrac{A}{m'}\right)}\cdot t\right)\right\} \quad \ldots . (18)$$

where C_k is a constant, $m' = m - \alpha d^2/v^2$, and v is the dislocation velocity. When $m' > 0$ the solution represents an oscillatory motion in which the atoms vibrate about fixed positions. However, when $m' < 0$, *i.e.* $v < d\sqrt{(\alpha/m)}$, the displacement changes steadily from $u_k = 0$ to $u_k = d$ (or *vice versa*) as the time varies from $-\infty$ to $+\infty$, and this corresponds to the propagation of slip along the row. In the limiting case of $m' = 0$, v takes the value $v_0 = d\sqrt{(\alpha/m)}$, which is the velocity of sound or the velocity of propagation of longitudinal waves, along the row. Since the motion of the dislocation is defined by the condition $v < d\sqrt{(\alpha/m)}$, this means that $v < v_0$, *i.e.* the velocity of slip propaga-

tion is always less than that of sound. By further development of the theory it can be shown that the total energy of the system is

$$W = \frac{4d^2\alpha}{\pi}\sqrt{\left(\frac{A}{m\,(v_0^2 - v^2)}\right)} \qquad \ldots\ (19)$$

which becomes infinite as the velocity approaches that of sound.

The interesting possibility has been put forward recently by FRANK[40] that fast dislocations might be able to use their high kinetic energies to create new dislocations, and he has discussed some mechanisms whereby this process could occur. Consider the arrival of a dislocation at a free surface of a crystal. As the dislocation passes out of the crystal, the atoms sliding past each other at the end of the slip plane will tend to overshoot and create a dislocation of opposite sign. If a new dislocation is formed, it will start to travel along the slip plane towards the interior of the crystal, under the action of the external force, thus continuing the glide process started by the first dislocation. The success of the attempt depends upon there being enough energy in a suitably ordered form to allow the creation of a new dislocation. The strain field of the first dislocation is one source of energy, but this by itself cannot be sufficient, since some energy will be dissipated as elastic waves when the first dislocation arrives at the surface and the surface energy of the crystal will be increased with the production of a step in the surface at the end of the slip plane. Thus the overshooting process could not possibly occur with a slow dislocation; for a fast dislocation approaching the surface, however, a large and highly ordered kinetic energy is also available, and this together with the strain energy should provide enough surplus energy to cause a new dislocation to be formed. This second dislocation will in turn become a fast dislocation and, at the opposite surface of the crystal, cause the formation of a third dislocation. It appears, therefore, that fast glide on a slip plane will be self perpetuating, and it is attractive to consider that this may be the explanation of the rapid, intense glide, localized to narrow slip bands observed in metal crystals.

By the same considerations as above Frank deduces that, if two fast edge or screw dislocations of opposite sign collide in a glide plane, they must pass through each other or be reflected from each other, these two processes being, of course, indistinguishable. This raises the possibility of a mechanism whereby dislocations can multiply in the

interior of a crystal, which is illustrated in *Figure 14*. In diagram a we have to imagine that a fast dislocation is approaching an obstacle in the glide plane, the latter being represented by the plane of the paper. The dislocation is halted or slowed down in the vicinity of the obstacle (diagram b), and a loop is formed. If the rest of the dislocation is moving very fast, overshooting occurs as the loop closes, and a region of the

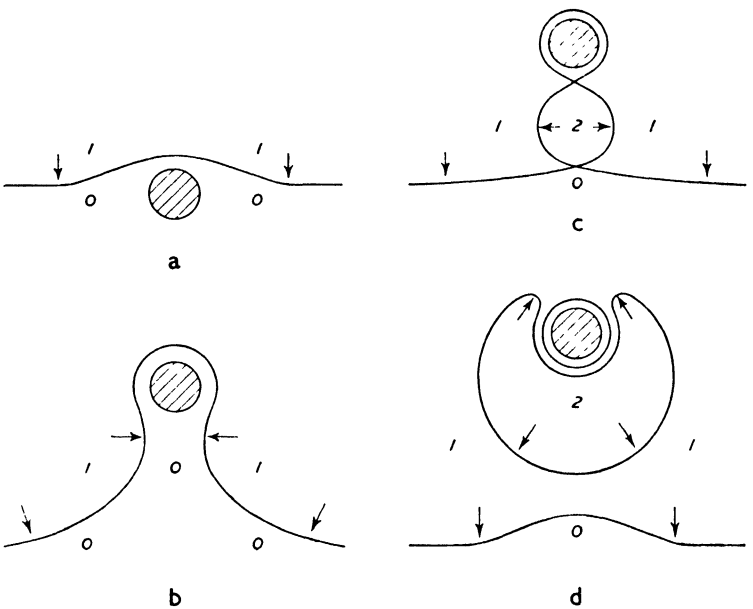


Figure 14. The multiplication of a fast dislocation by an obstacle. The numbers 0, 1, 2 in these diagrams show the amount of glide in different regions of the glide plane

glide plane with twice the amount of glide is formed (diagram c). This region then spreads and the process of glide multiplication continues.

The problem of why rapid glide on a slip plane eventually ceases is considered in Frank's theory in terms of the building up, in the plane, of a high density of dislocations, due to multiplication processes. It is argued that the final structure most likely to be developed in the glide plane is a crossed grid composed of two intersecting sets of parallel screw dislocations and also that, when the spacing of the grid becomes sufficiently small, the mutual interference of the dislocations will cause a 'seize-up' preventing further glide on that plane. This point is further discussed in the next section. Finally we may note that a crossed grid

of screw dislocations constitutes a transition surface between two crystals rotated with respect to each other about an axis normal to the transition surface. Experimental observation of such rotation of the crystal halves on either side of a slip band has been claimed by HEIDENREICH and SHOCKLEY.[41]

STRUCTURAL STRAIN HARDENING

As emphasized by MOTT and NABARRO,[14] it is necessary to consider strain hardening in terms of two distinct mechanisms, *structural* hardening and *exhaustion* hardening. Structural hardening is the well known type produced by cold working; exhaustion hardening is considered later.

The fact that the resistance to slip in a crystal increases with the amount of previous plastic deformation must mean either that it becomes more difficult to produce dislocations in the crystal or that it becomes more difficult to move them. It is usually assumed that the hardening is due to the difficulty of moving dislocations, and, as Mott and Nabarro[14] have pointed out, the experimental facts themselves support this suggestion, since the internal stresses which are known to occur in strain hardened crystals ought to help the formation of dislocations but hinder their motion. Thus the starting point for most modern theories of structural strain hardening is the idea, introduced by TAYLOR,[2] that the yield strength is determined by internal stresses opposing the motion of dislocations, these stresses becoming larger as deformation proceeds. The limited slip which occurs on a given glide plane suggests that the process of slipping leaves in its wake a residual change in the structure of the plane, which is a source of internal stresses opposing further slip. On the other hand, a single dislocation passing through a crystal leaves no such residual change; we are thus led to assume, with Taylor, that, except for initial ones which serve as ancestors for an avalanche of dislocations, the dislocations do not pass completely through a crystal but in some way become 'stuck' within the crystal. It is these stuck dislocations which are the sources of the stresses opposing the motion of others.

The simplest arrangement which would cause dislocations to become stuck is that shown in *Figure 15*. Here two dislocations of opposite

sign on parallel slip planes are attempting to pass each other under the action of the external force; the cross represents the positive dislocation, moving to the right, and the circle is the negative dislocation. From equations 9, the component F_x along the slip plane of the force per unit length between the dislocations is:

$$F_x = -\frac{\partial V}{\partial x} = \frac{G\lambda^2}{2\pi(1-\nu)} \frac{h^2x - x^3}{(x^2+h^2)^2} \quad \ldots . \quad (20)$$

where x and h are defined as in *Figure 15*. The variation of F_x with x/h is given in *Figure 16*. We notice that for small x the dislocations repel

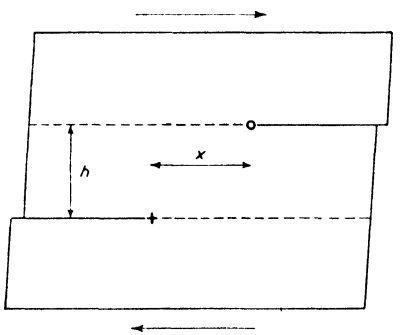


Figure 15. The formation of a pair of stuck dislocations

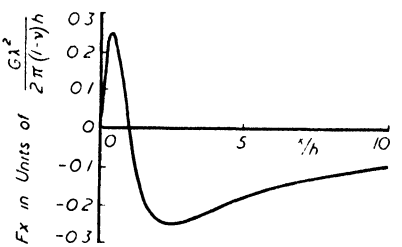


Figure 16. The force in the slip direction between the dislocations of Figure 15

each other along the slip planes and that the point $x = 0$ is one of unstable equilibrium with respect to the F_x force component; the strong attraction between dislocations of opposite sign is expressed by the F_y component, not F_x, when the dislocations are in this position. For larger values of x, F_x changes sign and the dislocations attract each other along the slip planes. It should be remembered that the interaction energy calculations on which *Figure 16* is based are approximate. The detailed form of the curve depends on the approximation; KOEHLER,[19] using a different approximation, obtains a qualitatively, but not quantitatively, similar curve.

Evidently, to drive the dislocations apart, the maximum value, $G\lambda^2/8\pi(1-\nu)h$, of F_x must be overcome, and this is equivalent to a shear stress:

$$\sigma = \frac{G\lambda}{8\pi(1-\nu)h} \quad \ldots . \quad (21)$$

Using the previous values for the constants, the numerical values of σ given in *Table III* are obtained. These results show that very large forces are required to drive dislocations on neighbouring planes past each other. If these values cannot be attained the dislocations must remain stuck, forming a knot which inhibits slip on all planes passing through or near it. The value of h corresponding to the yield strength of a fairly heavily strained crystal is about 10^{-6} cm. If dislocations are more openly spaced than this, yielding can still occur at this stress level.

Table III. The Shear Stress required to drive Dislocations on Neighbouring Slip Planes past each other

Distance between planes h (*cm*)	10^{-7}	10^{-6}	10^{-5}	10^{-4}
Shear stress (*gm mm*$^{-2}$)	60,000	6,000	600	60

Thus we may conclude that the density of dislocations in the cold worked state must be about 10^{12} cm^{-2}. On the same argument the density in a fully softened crystal could hardly exceed 10^{8} cm^{-2}, since otherwise the yield strength would be greater than is observed. The argument for a dislocation density of 10^{12} cm^{-2} for the cold worked state is so general that it is important to know whether this density is consistent with other properties. Thus several investigators[19, 42, 43] have compared the energy of the array of dislocations with the energy absorbed in cold working. TAYLOR and QUINNEY[44] have shown that about 43×10^{7} ergs per cc are stored in cold worked copper. If we take the energy of a dislocation as 5 electron volts per atom plane, then, in a one cm cube of copper containing 10^{12} dislocation lines, the energy of the dislocation is 32×10^{7} ergs, which is of the correct order of magnitude. Further evidence has been provided by BROWN,[20] who by interpreting the magnetic hardness of a cold worked ferromagnetic

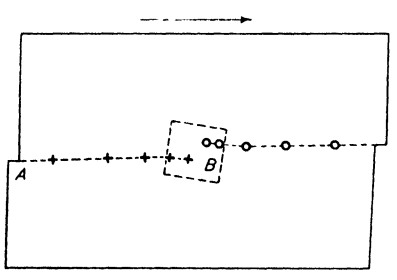


Figure 17. Pressure of dislocations at an obstacle producing local stresses and lattice rotation

metal in terms of an array of stuck dislocations, has shown that the experimental observations are consistent with a density of 10^{12} cm^{-2}.

Returning to the discussion of *Figure 15*, we observe that partial slip can still occur on the planes passing through the region of the knot, but the dislocations will not be able to break through until a high local stress is built up; thus an arrangement of the type of *Figure 17* can form. Three points should be noticed in connection with this diagram. In the first place, all dislocations which approach an obstacle from the same side have the same sign. Secondly, the density of dislocations must increase as the irregularity is approached, due to the pressure of dislocations along the plane. The pressure P exerted on an obstacle by a row of n dislocations of the same sign, the nth member of which is held fixed by the obstacle, can be calculated as follows: let the forces acting on the ith dislocation from the kth dislocation, the external stress σ and the obstacle be F_{ik}, $\lambda\sigma$ and P_i respectively. Then $P = -\Sigma P_i$; also if $i > k$ then F_{ik} acts in the direction of $\lambda\sigma$, *i.e.* against P_i and *vice versa*. The equilibrium of the ith dislocation under its forces is given by:

$$\lambda\sigma + \sum_{k=1}^{k=n} F_{ik} + P_i = 0$$

If we sum this expression over all dislocations and observe that

$$\sum_{i=1}^{i=n} \sum_{k=1}^{k=n} F_{ik} = 0, \text{ since } F_{ik} = -F_{ki},$$

we obtain:

$$P = n\lambda\sigma \qquad \ldots . \quad (22)$$

The pressure of dislocations thus produces a high local stress in the vicinity of the obstacle of n times the external stress. The third point to notice is that the dislocations will tend to rotate the lattice, as shown in *Figure 17*. BURGERS[12] has suggested that this may be the explanation of the crystal fragmentation and lattice rotation which x-ray investigations have shown to be characteristic of the cold worked state.

The broad picture emerging from the above arguments is that, as deformation proceeds, the density of dislocations steadily increases causing higher interaction forces and yield strength. To make a definite model of this effect, capable of quantitative treatment, Taylor[2] considered an idealized work hardened structure containing a regular array of dislocations, one form of which is shown in *Figure 18*. The

properties of such arrays have been further analysed by Koehler.[19] Taylor assumed in his treatment that the ratio a/b of the spacings of the array remained constant during straining, and with this he was able to account for parabolic stress/strain curves observed with metals of cubic crystal structure. From equation 21 it is clear that the stress required to shear the planes of dislocations past each other is of the form $\sigma = K/a$, where $K = G\lambda k/8\pi(1-\nu)$ and k is a constant whose precise value depends on the exact form of the dislocation array; since the interaction between nearest neighbours must be the dominating effect, k will be of the order of unity. To calculate the strain we remember that $\gamma = \rho l\lambda$, where $\rho = 1/ab$ is the dislocation density and l is the average distance travelled by a dislocation. If L is the maximum distance a dislocation can travel before becoming stuck, then $l = L/2$ and the shear strain is $\gamma = \lambda L/2ab$. Taking $a = b$, and combining the expressions for σ and γ, we obtain:

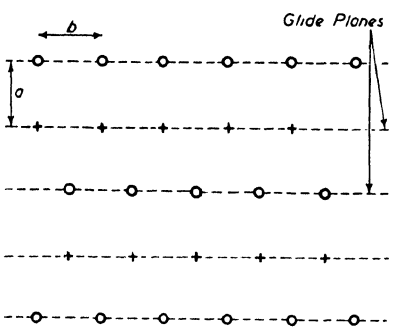


Figure 18. An array of dislocations (After Taylor)

$$\sigma^2 = \frac{2K^2}{\lambda L}\gamma \qquad \ldots\ (23)$$

which is the parabolic stress/strain relation.

The presence of the length L in this relation is rather strange. It might be supposed that L is the width of the crystal, but this cannot be correct since it would mean that the stress/strain curve depended upon the crystal size which is contrary to experimental observations. The order of magnitude of L can be determined from the relation $\gamma = \rho L\lambda/2$; if we take $\gamma = 1$ when $\rho = 10^{12}$, then L is about 10^{-4} cm. The consistency with which this value for L was obtained on several metals led TAYLOR[25] to suggest that the sources of obstruction to the dislocations are the boundaries between mosaic blocks, since these are known to be about 10^{-4} cm apart. An important assumption affecting L is that the dislocations are uniformly spaced in the material. SEITZ and READ[42] have pointed out that the plastic strain of a crystal is confined

to slip bands, and if equation 23 is to be applied to the material within the bands, then L must be greatly increased to correspond to the high local values of the strain. Further doubt on Taylor's interpretation of L has been raised by OROWAN,[45] who has pointed out that the observed similarity in the plastic properties of rock salt crystals of widely different degrees of perfection, as indicated by x-rays, does not support the mosaic boundary explanation. Orowan suggests that the obstructions may be dislocation knots of the type shown in *Figures 15* and *17*.

Another point of doubt concerning Taylor's theory is that hexagonal metals obey a linear hardening law instead of the parabolic law of equation 23. To obtain a linear relation between σ and γ the dislocation spacing b, in the slip direction, would have to be substantially constant, which is not easy to understand. As ANDRADE and ROSCOE[46] have pointed out, it is possible that all metals obey a linear hardening law. It is characteristic of the hexagonal metals that the number of glide bands is constant during plastic flow, whereas in aluminium new glide bands are formed continually, the spacing between them being roughly proportional to the reciprocal of the stress. If a fundamental linear hardening law for the material within the glide bands is assumed, it is easy to show that this would give a linear stress/strain curve for hexagonal metals and a parabolic one for aluminium.

A rather different picture of strain hardening has been introduced by BURGERS[47] and developed by KOCHENDÖRFER[48] and LAURENT.[49] Dislocations are considered to form by the combined action of stress and thermal agitation at places where a notch effect exists and causes a local stress concentration; crystallite boundaries are suggested as the most likely places. The dislocations created at these places run into the crystal until they meet obstructions, where they accumulate and, according to Burgers, cause crystallite rotation. It is also suggested that these dislocations collectively produce a stress field which counteracts the local stress at the notch, thus making the production of further dislocations more difficult; in this way an explanation of strain hardening is built up. Kochendörfer has extended these ideas by supposing that crystallite boundaries form the obstacles to dislocations in addition to acting as the sources of dislocations and that, by means of thermal agitation, a dislocation held up at a boundary may pass into it and

become annihilated. In this way it becomes possible to introduce dynamic effects into the theory and to discuss the effects of temperature and rate of straining upon the stress/strain curve. While the flow equations developed from these ideas can be made to fit several experimental facts, the rather large number of assumptions introduced prevents the theory from being completely convincing.

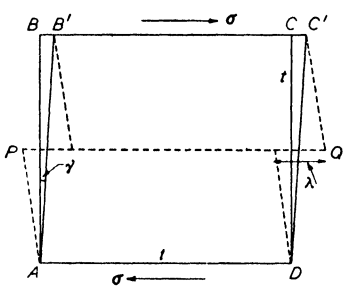


Figure 19. Slip in a crystallite (After Bragg)

A third approach to the problem of strain hardening has been made by BRAGG[23,43,50] from a consideration of whether the strain energy in a crystal is decreased or increased by slipping.

Consider the square crystal *ABCD* in *Figure 19*, with sides t, sheared an amount γ by a stress σ; the crystal is fixed in this position *AB' C' D* by external constraints at *A*, *B'*, *C'*, and *D*. Then, if slip by the smallest possible amount λ occurs along *PQ*, the configuration of the crystal becomes that given by the broken line. Bragg inquires whether the strain energy of the crystal after this unit of slip has occurred is higher or lower than before; if it is higher, the slip is assumed to be unstable and not to occur. Slip is assumed to occur if the strain energy is lowered by the process, thus implying that the large activation energy for starting slip is in some way overcome. It is clear from *Figure 19* that if the shear γ is small, *i.e.* $BB' << \lambda$, the crystal will be more highly strained after slip than before and the slip will not be stable. When $BB' \geqslant \lambda/2$ the strain energy in the crystal is reduced. The criterion of stability is therefore $BB' = \lambda/2$, and since $BB' = \gamma t$ and $\sigma = G\gamma$, the corresponding critical shear stress is:

$$\sigma = G\lambda/2t \qquad \ldots . \quad (24)$$

If we tentatively identify this criterion for strain energy reduction with the criterion for slip, equation 24 should define the yield strength. The important quantity is t, the crystal size. x-Ray evidence[51] suggests that t is about 10^{-5} cm in heavily worked metals of high melting point. Taking the previous values for G and λ, this gives $\sigma \simeq 5\,\mathrm{kg/mm^{-2}}$,

which is rather low for a heavily worked metal but is of correct order of magnitude.

As Bragg[23] has pointed out, this simple treatment neglects the fact that a crystallite embedded in a matrix of other crystallites must conform to its neighbours. This means that the crystallite is not in fact held along *DA* and *B'C'* by rigid constraints, since it is mounted in an elastic and therefore deformable medium. Also, the localized slip must create two misfits, *P* and *Q*, at the ends of the slipped plane; these are, of course, dislocations. Reconsideration of the problem from this point of view soon leads to the conclusion that the only part played by the boundary is to limit the slipped region to an interval of distance t in the glide direction, and that, excepting the misfits at *P* and *Q*, conditions along the crystallite boundary do not enter the problem since, in general, the elastic properties of the medium on either side are the same. In two dimensions the problem is thus that of the generation in a crystallite of a dislocation pair, the members of which move in opposite directions along the slip plane until they are halted at the crystallite boundaries. The problem of the stability of a dislocation pair in the same glide plane has already been considered previously, where it was shown that glide becomes stable when a critical distance of separation of the dislocations is exceeded, this distance being defined by a balance between the external force and the attraction between the dislocations. From *Table I* we see that for $t = 10^{-5}$ cm the minimum shear stress for stable glide is $\sigma = 2{\cdot}5$ kg/mm^{-2}, which is not greatly different from that given by the simple theory.* We have here used a different criterion of stable glide from that of Bragg;[23] if it is assumed that the glide is not stable until the energy at the end of the slip process is lower than that at the start, then the minimum shear stress is raised by a factor of about ten.

One problem facing the theory is the question of how the large activation energy for nucleating slip can be acquired, since thermal fluctuations are not strong enough. One is forced to assume that slip must start at weak spots where it is not necessary to supply the large energy required for creating dislocations in a perfect lattice. Bragg

* It should be noted that this is not just an accidental numerical agreement. Both equation 24 and the equation from which *Table I* was constructed give the yield as a small multiple of $G\lambda/t$.

suggests that these weak spots may be dislocations attached to crystallite boundaries. This problem is, however, one which faces the general theory of dislocations and will be discussed later.

Bragg has also proposed a further connection between crystallites and dislocations in worked metals.[43] Assuming that the dislocations introduced by cold working are all assembled in transition surfaces between crystallites and that the crystallite size is 10^{-5} cm, it is shown that there is one dislocation every 2·5 atomic spacings along the transition surfaces. In a two-dimensional model this gives an orientation difference of 30° between neighbouring crystallites, which is the largest possible angular difference for close packed crystal structures having 60° symmetry. If, therefore, there is a lower limiting crystallite size of 10^{-5} cm, as suggested by Wood,[51,52] the upper limits to the density of dislocations and the energy of cold working can be explained, since the transition surfaces become as crowded as possible with dislocations. However, Taylor's argument, which explains the strength of strain-hardened metals in terms of the closeness of approach of dislocations, and which leads to a density of 10^{12} cm^{-2} independently of energy considerations, suggests that the dislocations are at least quasi-regularly packed and are not all assembled in transition surfaces.

AGEING AND ANNEALING OF WORKED METALS

On the picture developed in the previous section, a cold worked metal is in a thermodynamically unstable state because it contains a large number of dislocations. Thus, for a consistent theory, it is necessary to explain the spontaneous changes which occur on annealing after cold work in terms of processes associated with these dislocations. Two such processes suggest themselves, each leading to a reduction of free energy, *i* the segregation of solute atoms round dislocations (already discussed) and *ii* the migration and dissolution of dislocations. In practice three changes, strain ageing, recovery, and recrystallization, may be observed on annealing a worked metal, although in many instances one or both of the first two may either be absent or be obscured by recrystallization.

Strain ageing is the name given to the increase in hardness sometimes produced by annealing at temperatures below the range where recovery and recrystallization become rapid. In iron the process is associated

with the return of the yield point and was explained earlier by the segregation of carbon atoms round dislocations. A very similar effect, thermal hardening, discovered by Orowan[37] in zinc crystals, and also observed in cadmium,[53] may meet with the same explanation. Several other effects occur on low temperature annealing which appear to be different aspects of the same general phenomenon. A freshly worked metal shows certain 'anelastic' effects,[54] of which high internal friction[55] and slight plasticity at stresses well below the normal yield strength[56] are the most prominent. These have been attributed by ZENER[54] to the movement of dislocations in the slip planes of the worked material. This is reasonable, since local readjustments of free dislocations ought to occur under quite small stresses, even though a high stress is required to overcome the interaction forces of neighbouring dislocations and so produce extensive plastic flow. Low temperature annealing greatly reduces the internal friction and causes the return of a true elastic range; the explanation of these in terms of the anchoring of the dislocations by solute atom atmospheres is consistent with the observation of READ and TYNDALL[57] of the high sensitivity to impurities of the anelastic effects in zinc crystals.

At rather higher temperatures than are necessary for strain ageing a worked metal resoftens by the processes of recovery and recrystallization. Recovery becomes rapid at lower temperatures than recrystallization and produces a partial softening without any visible accompanying changes in the microstructure of the material. Recrystallization, on the other hand, can cause full softening and involves the nucleation of new, undistorted crystal grains which subsequently grow at the expense of surrounding distorted material and each other. Thus recrystallization is complete when the entire material is composed of new grains, although the sequence of changes associated with the annealing of a worked metal is usually not complete at this stage for grain growth can still occur. From the standpoint of the dislocation theory it is natural to interpret these processes of thermal softening in terms of the movement and dissolution of dislocations. As we have seen, some support for this view is given by CAHN's work on polygonization.[28] Also, the great sensitivity of thermal softening to the presence of small traces of impurity[58,59] is consistent with a dislocation explanation since small amounts of solute can, by atmosphere formation, greatly reduce the

mobility of dislocations. Thermal softening has been discussed in terms of dislocations by several writers[6,19,42] and considered in some detail by Burgers.[60] Burgers starts from a picture of the cold worked state which is similar to that proposed by Bragg *i.e.* an assembly of crystallites, elastically deformed by internal stresses, which are joined by transition surfaces composed of dislocations. It is, however, admitted that some dislocations may exist in the interiors of the crystallites. In addition to the simple transition surfaces of the types discussed earlier, Burgers also considers more complicated ones, of several atoms width, in which dislocation arrays similar to that of *Figure 18* exist. In such cases the transition layers contain dislocations of opposite sign and an orientation difference between neighbouring crystallites necessitates an excess of dislocations of one sign.

Two softening processes are discussed. In the first, which is considered to account for recovery, a reduction in the number of dislocations in the boundaries occurs without any displacement of the boundaries as a whole, or any rotation of neighbouring crystallites relative to each other. The latter condition necessitates that the dislocations be removed from a boundary in pairs, of opposite sign, and Burgers suggests two ways in which this might be accomplished. If dislocations of opposite sign exist on the same glide plane they can run together and annihilate each other. If they are not on the same plane they might, under favourable circumstances, be driven apart along their respective glide planes into the adjoining crystallites under the action of a local internal stress which could arise as a consequence of the dissolution and re-arrangement of nearby dislocations. These mechanisms are rather difficult to understand when the dislocations are close together, since the interaction forces are then very large. One would thus expect dislocations on the same plane to annihilate each other almost instantaneously, without waiting for thermal agitation; those on nearby planes would form a knot in which they are bound too firmly to be dislodged by the internal stress. Burgers points out, however, that other nearby dislocations may markedly affect these tendencies. A possible mechanism for the dissolution of a knot results from the fact that, geometrically, the insertion or removal of a part plane of atoms joining the two dislocations produces a perfect crystal structure. If the dislocations are arranged with their compressed sides innermost,

the part plane has to be removed, and *vice versa*. Since the dislocations in a tightly bound knot are, at most, only a few atoms apart, it ought to be possible to achieve this result in a reasonably short time by the transport of atoms to or from the region between the dislocations by migration. A process which could cause recovery by the dissolution of single dislocations has been suggested by Koehler.[19] He has shown that a necessary consequence of elasticity theory is that a dislocation near a free surface must experience an image force equivalent to the force which would be produced by a dislocation of opposite sign at an equal distance on the opposite side of the surface. Under the action of this image force dislocations can migrate to a free surface and become annihilated.

Whatever mechanism causes the dissolution of dislocation pairs, it cannot produce a complete removal of the effects of cold working since it must cease when each transition surface consists entirely of dislocations of one sign and no dislocations are present in the interiors of the crystallites. Recovery is considered to be completed at this stage, and further change must proceed by a different process. The characteristic feature of recrystallization is crystal growth, and Burgers accounts for this by invoking a mechanism due to Bragg[26] in which a transition surface migrates as a whole by the simultaneous movement of all its constituent dislocations along their respective glide planes. This process enlarges one crystallite at the expense of another and so produces crystal growth. It can occur when the resultant internal stress acting in the slip direction on the set of dislocations attains a certain value. To obtain a systematic migration of a transition surface it is necessary for the state of stress in adjoining crystallites to be different, and it is rather difficult to see how this condition is preserved during the continued growth of a crystal. Conditions which might allow a crystallite to act as a nucleus for recrystallization, as envisaged by Burgers, are illustrated in *Figure 20*. In this, the central crystallite occurs at the point of inflexion of a curved lattice region and is

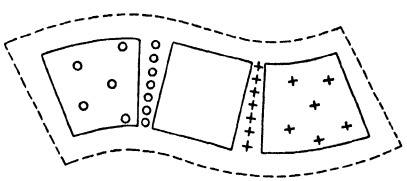


Figure 20. Adjoining crystallites in a distorted crystal. The central crystallite is free from internal stress and may serve as a recrystallization nucleus (After Burgers)

comparatively free from internal stress. As it is less deformed than its neighbours it may be expected to grow at their expense.

One phenomenon which lends some support to these ideas is that of stimulated crystal growth, reported by SANDEE, MAY, and BURGERS.[61] Analysis of the shapes of crystals grown in recrystallized aluminium plates showed that certain crystals did not start to grow until other crystals, already growing, reached their nuclei and stimulated them into activity. This effect appears to explain the observation of ANDERSON and MEHL[62] that the rate of nucleation in recrystallizing aluminium increases with time. It is highly suggestive that the stimulating and stimulated crystals are always oriented so that they have in common a glide plane (111) and a glide direction [110] in this plane. Burgers suggests that, when a growing crystal reaches a crystallite with a parallel set of glide elements, dislocations held up in the crystallite are 'discharged' into the crystal, thus relieving the crystallite of internal stresses and allowing it to grow as in the process envisaged in *Figure 20.*

FLEXIBLE DISLOCATIONS AND PRECIPITATION HARDENING

It is well known that a freshly quenched alloy, in the form of a supersaturated solid solution, is soft and that it becomes harder when precipitation is allowed to occur. If ageing continues further the particle size of the precipitate becomes large and resoftening occurs (over-ageing). The yield strength of the alloy is thus a function of the state of dispersion of the particles in the material, as represented qualitatively by *Figure 21.* In this diagram Λ_0 is the mean distance between neighbouring particles in the freshly quenched alloy; the particles in this case are individual solute atoms and Λ_0 is about three atomic spacings for typical alloy compositions. Maximum hardening occurs at a critical dispersion size, Λ_{crit}, which has been shown experimentally[64] to be about 25 to 50 atomic spacings. On over-ageing the dispersion eventually becomes sufficiently coarse for individual particles to become microscopically visible, so that in the softening range Λ may be expected to be of the order of 1,000 spacings and greater.

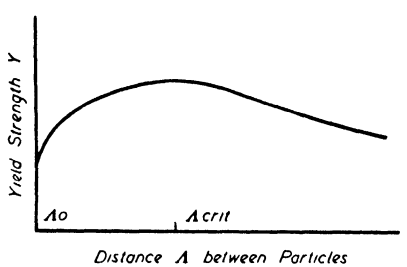


Figure 21. The variation of the yield strength of a precipitation hardening alloy with distance between particles

The dislocation theory of hardening due to precipitation has been built up by MOTT and NABARRO.[65,66,67,14] In their first paper[65] two ideas are introduced which form a basis for the subsequent development of the theory:

1 The misfit of particles in the solvent matrix causes internal stresses. In the solid solution this misfit is determined by the parameter ε, which measures the difference in the atomic radii of the solvent and solute, r_a and $r_a(1 + \varepsilon)$ respectively. On precipitation the solute atoms cluster together to form larger particles, but the degree of misfit is unaltered since a spherical region enclosing N atomic sites changes its radius from R to $R(1 + \varepsilon)$ when the solvent atoms in these sites are replaced by solute atoms.

2 The cause of hardening is the resistance offered by these internal stresses to the passage of dislocations. A moving dislocation must encounter lattice regions where the local stress acts against the external stress. To continue the motion the internal stress must be overcome, and it is assumed that the criterion for this is that the external stress should attain the same magnitude as the mean internal stress.

It is shown by Mott and Nabarro that this condition for yielding defines the yield strength in order of magnitude as

$$\Upsilon = E\varepsilon f \qquad \ldots . \quad (25)$$

where E is Young's modulus for the matrix and f is the ratio of the total volume of particles to that of the matrix. The important feature of this result is that it predicts the yield strength to be independent of Λ, the distance between particles, and to depend only on f, the total quantity of material causing internal stresses. This is because the degree of misfit is unaltered during precipitation. Elasticity theory calculations show that the intensity of the internal stresses is dependent upon ε but not upon Λ, so that precipitation increases the wavelength of the stresses but not their amplitude.

We might thus expect the solute atoms to exert their full hardening effect for all states of dispersion in the alloy. However, another line of reasoning suggests that the internal stresses could have almost no hardening effect. If we consider a dislocation as a rigid, linear discontinuity of length L, where $L >> \Lambda$, then along its length there will

be L/Λ randomly arranged local fields of internal stress, some of which help the forward motion of the dislocation and others hinder it. To a first approximation these randomly alternating forces acting on the dislocation cancel, so that their net effect is zero and there is no hardening.

The resolution of these difficulties and the key to the problem lie in the important idea, introduced later,[66,67] that a dislocation is not a rigid discontinuity but is flexible. From a macroscopic point of view, *i.e.* compared with atomic orders of size, a long dislocation can be considered as a smoothly curved discontinuity lying in the slip plane, the curvature being caused by the stress fields. Consider an element of length l of such a dislocation in a region where the local stress is σ_i. The force exerted on it by the stress is $\sigma_i \lambda l$, which means that the element has a potential energy $V(x) = \int \sigma_i \lambda l dx$ due to the stress, where x is the position of the element along the slip direction relative to the local stress field. It is thus possible to plot the potential energy for a dislocation at all points in a slip plane, and the resulting diagram is a contour map of randomly arranged potential hills and valleys in the slip plane, corresponding to the local stress fields, as shown in *Figure 22*. In this diagram the circles enclose regions in which the dislocation has high energy. In a stable configuration such regions are avoided and the dislocation follows a path along the valleys shown by the broken

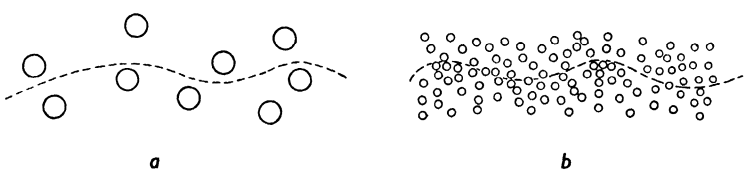


Figure 22. Flexible dislocations in internal stress fields. The slip plane is the plane of the paper. The dislocations are marked by broken lines and the circles enclose regions of high potential energy

line. If the dislocation were ideally flexible it would always curl completely round the energy hills. Actually this is not so, and if the hills and valleys alternate on a sufficiently fine scale the dislocation cannot follow the sharply curving contours and must adopt a form of the kind shown in *Figure 22b*. The reason for this is that the dislocation becomes longer as it develops more loops, so that its energy due to its stress field increases. Mott and Nabarro[14] have shown that the energy of a dislocation is about Gd^3 per atom plane, where G is the

shear modulus and d is the atomic spacing; this is of the order of a few electron volts. As a result a dislocation may be thought of as having a tension

$$T = Gd^2 \qquad \ldots\ (26)$$

acting along its length trying to shorten it, and is thus analogous in this respect to a one-dimensional soap film. It can be shown that the shape adopted by a flexible dislocation in a field of internal stress is defined by

$$T/\rho = \sigma_i d \qquad \ldots\ (27)$$

where ρ is the local value of the radius of curvature of the dislocation line in the region where the internal stress is σ_i. Eliminating T gives

$$\rho = Gd/\sigma_i \qquad \ldots\ (28)$$

so that the radius of curvature in atomic spacings is equal to the ratio of the shear modulus to the internal stress.

We now have to compare ρ with Λ, the spacing between the centres of strain. It is clear that, if $\Lambda > \rho$, the dislocation can follow the contours of the stress field and a configuration like that of *Figure 22a* can be adopted. In such a case the dislocation is looped into sections of wavelength and amplitude of the order of Λ and, because it is quite flexible at this radius of curvature, each loop will act more or less independently of the others. Hence to produce glide each loop will have to be taken over a potential hill without help from other sections of the dislocation. To accomplish this the external stress will have to exceed σ_i and the alloy will be hard (the over-ageing case where $\Lambda >> \rho$ is discussed below). If $\Lambda < \rho$, on the other hand, the dislocation cannot follow all the changes in the stress field and must appear as in *Figure 22b*. This is an approach to the case of a rigid dislocation, for which the internal stresses balance and cause no hardening. Clearly this example is to be identified with that of the freshly quenched solid solution, in which Λ is only a few atomic spacings. To obtain a dislocation curvature of this magnitude the internal stresses would have to be of the same order as the shear modulus, whereas in practice the stresses round a solute atom cannot reasonably exceed $G/100$.

We thus have a basis for understanding why an alloy hardens as it changes from the condition represented by *Figure 22b* to that of *Figure 22a*. This hardening is expected to become complete when the particles

have grown to such a size that $\Lambda = \rho$, which ought to be the critical dispersion size, Λ_{crit}. The estimation of this quantity involves a knowledge of σ_i. From the theory we know that σ_i should be about the same as the yield strength Υ for a fully hardened alloy. Experimentally Υ is about $G/100$ in such cases, from which we deduce that Λ_{crit} should be about 100 atomic spacings. This is in fair agreement with Guinier's experimental value.[64]

According to the theory, the yield strength increases proportionately with $\Lambda^{1/2}$ as the alloy passes from the as-quenched to the fully hardened state. Thus, consider a dislocation of length $L >> \Lambda$. This is divided into L/Λ elements by the internal stress field, in each of which the stress is σ_i and the sign of σ_i varies randomly from one element to the next. Hence the dislocation is subjected to L/Λ randomly positive and negative forces, each of magnitude $\sigma_i \lambda \Lambda$. From statistical theory it is well known that the probable net value of n randomly positive and negative equal contributions is $n^{1/2}$ times the value of a single contribution. The net force on the length L is therefore, on average, $\sigma_i \lambda \Lambda (L/\Lambda)^{1/2}$, which is equivalent to an average shear stress $\sigma_i (\Lambda/L)^{1/2}$. The resultant internal stress (and hence the yield strength) thus increases as $\Lambda^{1/2}$.

The mechanism of resoftening on over-ageing has been discussed by Orowan,[68] who has shown that it can be explained as a necessary consequence of the concept of flexible dislocations. Consider in *Figure 23* a dislocation line held up by widely spaced regions of high potential energy. Under the action of the external stress the dislocation will bulge forward into the regions between the obstacles (line A), and this bulging will increase as the applied stress increases (B). Eventually the dislocation will break away from the obstacles (C), leaving them enclosed by small dislocation loops. The yield stress is thus that required to push a dislocation line through a row of obstacles of spacing Λ, and to achieve this the dislocation must be forced into loops whose radii of curvature are approximately $\Lambda/2$. The shear stress required to bend a dislocation to such a curvature is given by equation 28 as

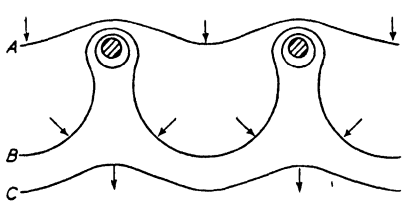


Figure 23. The by-passing of widely spaced obstacles by dislocations (After Orowan)

$$\Upsilon = 2Gd/\Lambda \quad \ldots . \quad (29)$$

so that, in the over-ageing range, the yield strength should decrease inversely with Λ.

An additional effect which ought to be taken into account arises from the pressure of the dislocations which are held up by the obstacles. We have already seen that a dislocation which is held in equilibrium against an obstacle by the pressure of $n - 1$ other dislocations behind it is subjected to balanced forces equivalent to a local stress which is n times the value of the external stress. We might expect from this that, by forming sufficient dislocations in a glide plane, any given obstacle can be overcome with an arbitrarily small external force. This is only true if there is an unlimited expanse of glide plane behind the obstacle in which the dislocations can be accommodated. In practice fresh obstacles occur at intervals of Λ in the plane, and the pressure exerted on a dislocation next to an obstacle depends upon how many dislocations can be accommodated within such an interval. As Λ increases, more dislocations can be accommodated, which means that a greater pressure can be exerted on a dislocation forcing it through the region between the obstacles. Thus increasing Λ can not only make it easier for dislocations to by-pass obstacles but can also permit a greater pressure, for a given external stress, forcing them through.

Mott and Nabarro have recently extended their theory to take into account the effect of thermal fluctuations in helping dislocation loops over the potential energy hills in the slip plane.[14] In the fully hardened alloy a loop of length Λ will, in the absence of an external stress, have to surmount a potential barrier $\int \sigma_i(x)\lambda\Lambda dx$, where $\sigma_i(x)$ is the mean stress along the loop when the latter is displaced a distance x from the valley position. The distance the loop has to travel before reaching the next valley is of order Λ, and, assuming that the internal stress varies sinusoidally across this region, we can take $\sigma_i(x) = \sigma_i \sin(2\pi x/\Lambda)$. Integrating with respect to x gives the potential energy hill as

$$\sigma_i \lambda \Lambda^2/\pi \qquad \ldots . \quad (30)$$

If we assume that $\sigma_i = G/100$, $\Lambda = 100d$, $\lambda = d$, and $Gd^3 = 5$ev, then the energy barrier is 150 to 200 ev. There can be no chance of thermal fluctuations providing such an activation energy. They can only provide energies of about one electron volt and less with reasonable frequency at room temperature. It is clear that an external stress σ almost equal to σ_i would have to be applied before the barrier was

reduced to such a value, and this leads to the result that the yield strength ought to be almost independent of the testing temperature. While a specific experimental test of this prediction on precipitation hardened alloys does not seem to have yet been made, the fact that a large temperature dependence is observed in many annealed metals and single crystals is not encouraging. Mott and Nabarro conclude that some additional factor, such as a temperature dependence of σ_i, must be operative.

On the assumption that flow occurs when the external stress σ is almost equal to σ_i, the activation energy for moving a dislocation loop from one valley to the next is found to be:[14]

$$U(\sigma_i) = 0{\cdot}15\sigma_i d\Lambda^2\left(1 - \frac{\sigma}{\sigma_i}\right)^{3/2} \quad . \ . \ . \ . \quad (31)$$

and in precipitation hardened alloys, at least, this ought to replace the well known Becker-Orowan activation energy formula

$$U(\sigma_i) = \frac{v(\sigma_i - \sigma)^2}{2G} \quad . \ . \ . \ . \quad (32)$$

where v is the volume in which a thermal fluctuation causing glide takes place. Experimentally, a distinction between these formulae should be possible on the basis of their different dependences on σ; existing data are not, however, sufficiently precise for this purpose.

TRANSIENT CREEP

Apart from micro-creep, two main kinds of creep have been recognized, transient and quasi-viscous. Transient creep occurs when a metal is loaded beyond its yield strength. After an initial instantaneous extension, a loaded specimen continues to extend under constant stress, the plastic strain γ increasing according to the law

$$\gamma = \beta t^{1/3} \quad . \ . \ . \ . \quad (33)$$

where t is the time and β is a coefficient of flow. Initially the fast transient creep obscures other creep processes, but with increasing time it becomes slow and, in most cases, a second process is then revealed, occurring simultaneously and obeying the relation

$$\gamma = kt \quad . \ . \ . \ . \quad (34)$$

where k is a coefficient of flow. This is quasi-viscous creep. These laws were first established by ANDRADE[69] on polycrystalline metals and

have subsequently been confirmed on zinc[70] and aluminium[71] crystals. Several experimental facts show that the two flows are quite different. Thus quasi-viscous creep is strongly dependent on temperature and vanishes at low temperatures, whereas transient creep can be quite rapid at extremely low temperatures.

Quasi-viscous creep is not yet fully understood,[42, 71] and it is probable that different mechanisms operate in different cases. In polycrystalline metals it is usually caused by movements at or near the grain boundaries[72] and in single crystals is probably associated with movements in slip planes where previous flow has caused some atomic disorder.[72] SEITZ and READ[42] and KAUZMANN[73] have discussed some possible dislocation processes which could give quasi-viscous creep.

Interest has centred recently in transient creep and the elements of a theory are beginning to appear, based on two important features of this type of flow. It is a fast crystallographic flow, intimately connected with the instantaneous plastic deformation occurring on loading, and is thus almost certainly produced by fast dislocations. The fact that the initial flow rate is very fast, even at the lowest testing temperatures, implies that the activation energy required to move dislocations is vanishingly small at the start of transient creep.

The chance of a dislocation loop jumping through a region of unfavourable stress in unit time can be written as

$$\alpha = \nu e^{-U(\sigma_i)/kT} \qquad \ldots . \quad (35)$$

where $U(\sigma_i)$, the activation energy for the jump, is given by either equation 31 or 32, and ν is the frequency of vibration of the dislocation in the potential valley. MOTT and NABARRO[14] have shown that ν is about 10^8 sec^{-1}. The first interpretation of transient creep from this standpoint was given by OROWAN[71] and based on the assumption that, below the softening range of temperature, the plastic behaviour of a metal is determined by a stress/strain curve which is the expression of structural strain hardening, *i.e.* a progressive increase of the internal stress σ_i with strain. Thus, in *Figure 24*, application of a stress σ produces the instantaneous strain OA and, at the point P, the external and internal stresses, σ and σ_i, are equal. At 0°K no further flow would then occur, since $\sigma_i > \sigma$ beyond this point, but at higher temperatures thermal stress fluctuations can span the difference $\Delta\sigma$ between σ_i and

σ, thus allowing flow to continue at constant stress. At first, since $\Delta\sigma$ is vanishingly small, very small fluctuations are required and so rapid flow will occur even at very low temperatures. This continuing strain causes $\Delta\sigma$ to increase, however, and successive fluctuations have to span an ever widening gap between σ_i and σ. Successful fluctuations thus occur less frequently and the creep slows down. If we write $\Delta\sigma = h\gamma$, where the creep strain γ is measured from the point A in *Figure 24*, and assume that each fluctuation produces the same increment of strain, then the creep rate can be shown to be:[71,14]

$$\frac{d\gamma}{dt} = \text{const.}\, e^{-vh^2\gamma^2/2GkT} \qquad \ldots\ (36)$$

when the Becker-Orowan formula (equation 32) is used, or alternatively, an analogous expression with $\gamma^{3/2}$ replacing γ^2 if Mott and Nabarro's (equation 31) formula is used. Equation 36 gives a flow which starts from P at a finite rate only, and this is inconsistent with the fact that there is a continuous transition from the instantaneous extension to the transient flow. To overcome this difficulty Orowan assumed that, when $\Delta\sigma$ is small, the movement of one dislocation sets others into motion, producing a total increment of glide proportional to $1/\gamma^2$. With this assumption he obtained the creep formula

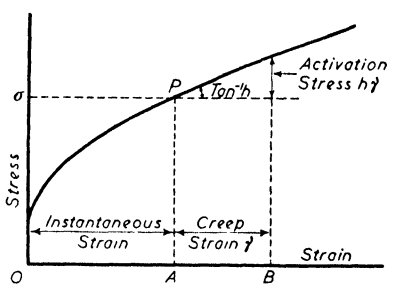


Figure 24. Interpretation of transient creep from a stress/strain curve (After Orowan)

$$\frac{d\gamma}{dt} = \frac{\text{const}}{\gamma^2} e^{-vh^2\gamma^2/2GkT} \qquad \ldots\ (37)$$

At the beginning of creep this can be approximated to $d\gamma/dt = \text{const}/\gamma^2$, the integral of which gives Andrade's $\gamma = \beta t^{1/3}$ formula.

It is assumed in Orowan's theory that, at any point on the stress/strain curve, only a single value of σ_i is operative and that structural hardening occurs during transient creep *i.e.* σ_i increases with γ. MOTT and NABARRO [14] and SMITH [74] have developed a rather different picture in which the assumption of a single σ_i is not made. If a range of internal stresses is present, then at any instant some dislocations will be in positions where σ_i is small and can easily be moved, whereas

others will face large stress barriers and their movement will require large thermal fluctuations. The lightly bound dislocations will evidently move first, and, if these continue to move until halted at places where σ_i is high, the creep will gradually slow down because of the exhaustion of lightly bound dislocations. This mechanism thus gives a decreasing creep rate without invoking an increase in σ_i, *i.e.* without invoking structural strain hardening, and accordingly it has been called exhaustion hardening. If it is assumed that no structural hardening occurs during transient creep it can be shown that exhaustion hardening should lead to a strain/time relation of the type:

$$\gamma = \text{const}\,\{\log(\nu t)\}^{x} \qquad \ldots\ (38)$$

where ν is about 10^8 sec^{-1} and x varies with the details of the theory and is given as 2/3 by Mott and Nabarro and as unity by Smith. This gives a creep curve rather similar to the $t^{1/3}$ curve and predicts a fast initial creep rate without involving the assumption that the movement of one dislocation starts a glide process producing an increment proportional to $1/\gamma^2$.

It is characteristic of the exhaustion theory that the extension produced in the first few seconds ought to be large compared with that produced later. Thus for $t = 1$ second the factor $\{\log(\nu t)\}^{2/3}$ is about 7, whereas for $t = 10^6$ seconds, approximately ten days, it is only about 10. Since all flow occurring in the first second would be regarded as instantaneous, the instantaneous strain ought always to exceed the transient strain for reasonable testing times. In practice the opposite is usually observed, and thus the concept of exhaustion hardening cannot by itself entirely account for transient creep. It is evident that an additional factor to be taken into account is structural hardening, and it is thus possible that a combination of Orowan's theory with the theory of exhaustion hardening will provide a more complete picture.

THE ORIGIN OF DISLOCATIONS

The excuse for delaying the discussion of the origin of dislocations until last is that this subject is at present the most obscure part of the theory of dislocations. Their high energy of formation makes it clear that dislocations cannot exist as stable features of a crystal in equilibrium and could hardly be formed by thermal fluctuations in even a highly

stressed but otherwise perfect crystal. The problem of the origin of dislocations is thus to explain how this high energy is overcome. That we have as yet no certain answer is no objection to the concept of dislocations and to the belief in their existence. The argument that dislocations must be present as a consequence of the crystallographic nature of plastic flow is convincing, and in any case, if we refused to accept it by supposing that glide planes moved rigidly across each other, we should again have the same problem to face, except that the energy to be overcome would be even higher.

Explanation has centred round the possibilities that dislocations originate either at the base of cracks where high stress concentrations exist[75] or at crystallite boundaries,[48,42,66,23] regarding these as dislocation transition surfaces. One difficulty facing the crack hypothesis is that extremely deep cracks, of the order of one mm, seem to be required to obtain sufficiently intense stress concentrations to produce dislocations at the observed yield strengths of soft single crystals.[76] Also one would expect plastic flow at the base of a crack to blunt it and impair its action in producing further flow. It has been suggested that the existence of slip bands supports the crack hypothesis rather than the other, since a large localized slip could be expected in the vicinity of a crack, whereas on the crystallite boundary hypothesis one would expect each slip plane to contribute a few dislocations, thus giving a homogeneous flow without visible slip bands. However, Frank's work on slip band formation by fast dislocations shows that slip bands may be produced without the aid of cracks, and accordingly recent opinion has tended to favour the crystallite boundary hypothesis.

Very recently Frank[39] has analysed the conditions governing the growth of a crystal about a nucleus from the vapour phase or the melt, and has shown that a nucleus is unlikely to grow unless it contains dislocations. If this is true it means that the problem of the origin of dislocations is essentially solved since dislocations will then be naturally occurring, though thermodynamically unstable, features of real crystals. Whether these dislocations are assembled in transition surfaces or not is a secondary question.

REFERENCES

[1] NADAI, A. *Plasticity* New York, 1931
[2] TAYLOR, G. I. *Proc. roy. Soc. A* 145 (1934) 362
[3] OROWAN, E. *Z. Phys.* 89 (1934) 634
[4] POLANYI, M. *Ibid* 89 (1934) 660
[5] PRANDTL, L. *Z. angew. Math. Phys.* 8 (1928) 85
[6] DEHLINGER, U. *Ann. Phys.* 2 (1929) 749
[7] EWING, A. and ROSENHAIN, W. *Phil. Trans. roy. Soc. A* 193 (1899) 353
[8] MARK, H., POLANYI, M. and SCHMID, E. *Z. Phys.* 12 (1927) 58
[9] TAYLOR, G. I. and ELAM, C. F. *Proc. roy. Soc. A* 102 (1923) 643
[10] ROSBAUD, P. and SCHMID, E. *Z. Phys.* 32 (1925) 197
[11] BURGERS, J. M. *Proc. roy. Acad. Sci. Amsterdam* 42 (1939) 293, 378
[12] —— *Proc. phys. Soc. Lond.* 52 (1940) 23
[13] BRAGG, W. L. and NYE, J. F. *Proc. roy. Soc. A* 190 (1947) 474
[14] MOTT, N. F. and NABARRO, F. R. N. *Rep. Conf. on Strength of Solids* Physical Soc. London, 1948 p 1
[15] TIMPE, A. *Z. Math. Phys.* 52 (1905) 348
[16] VOLTERRA, V. *Ann. Ec. Norm.* 24 (1907) 401
[17] LOVE, A. E. H. *A Treatise on the Mathematical Theory of Elasticity* Cambridge, 1934 p 221–8
[18] NABARRO, F. R. N. *Proc. phys. Soc. Lond.* 59 (1947) 256
[19] KOEHLER, J. S. *Phys. Rev.* 60 (1941) 397
[20] BROWN, W. F. *Ibid* 60 (1941) 139
[21] PEIERLS, R. *Proc. phys. Soc. Lond.* 52 (1940) 34
[22] COTTRELL, A. H. *Rep. Conf. on Strength of Solids* Physical Soc. London, 1948 p 30
[23] BRAGG, W. L. *Symposium on Internal Stresses* Institute of Metals 1947 p 221
[24] HUNTINGTON, H. B. *Phys. Rev.* 59 (1941) 942 *A*
[25] TAYLOR, G. I. *Proc. roy. Soc. A* 145 (1934) 388
[26] BRAGG, W. L. *Proc. phys. Soc. Lond.* 52 (1940) 54, 105
[27] LENNARD-JONES, J. E. *Ibid* 52 (1940) 38
[28] CAHN, R. W. *Rep. Conf. on Strength of Solids* Physical Soc. London, 1948 p 136
[29] NABARRO, F. R. N. *Ibid* p 38
[30] CHALMERS, B. *Proc. roy. Soc. A* 156 (1936) 427
[31] JASWON, M. A. and COTTRELL, A. H. To be published
[32] EDWARDS, C. A., PHILLIPS, D. L. and LIU, Y. H. *J. Iron Steel Inst.* 147 (1943) 145
[33] —— —— and JONES, H. N. *Ibid* 142 (1940) 199
[34] SNOEK, J. L. *Physica* 8 (1941) 734
[35] LOW, J. R. and GENSAMER, M. *Trans. Amer. Inst. min. met. Engrs* 158 (1944) 207
[36] COTTRELL, A. H. and BILBY, B. A. *Proc. phys. Soc. Lond.* 62 (1949) 19
[37] OROWAN, E. *Ibid* 52 (1940) 8
[38] FRENKEL, J. and KONTOROVA, T. *Phys. Z. Sowjetunion* 13 (1938) 1
[39] FRANK, F. C. Private Communication
[40] —— *Rep. Conf. on Strength of Solids* Physical Soc. London, 1948 p 46
[41] HEIDENREICH, R. D. and SHOCKLEY, W. *Ibid* p 57
[42] SEITZ, F. and READ, T. A. *J. Applied Phys.* 12 (1941) 100, 170, 470, 538
[43] BRAGG, W. L. *Trans. North East Coast Inst. Eng. Shipbuilders* 62 (1945) 25
[44] TAYLOR, G. I. and QUINNEY, H. *Proc. roy. Soc. A* 143 (1934) 307; *A* 163 (1937) 157
[45] OROWAN, E. *Symposium on Internal Stresses* Institute of Metals, London, 1947 p 47
[46] ANDRADE, E. N. DA C. and ROSCOE, R. *Proc. phys. Soc. Lond.* 49 (1937) 152
[47] BURGERS, W. G. and J. M. *Second Report on Viscosity and Plasticity, Acad. Sci. Amsterdam* 1938 p 200
[48] KOCHENDÖRFER, A. *Z. Phys.* 108 (1938) 244
[49] LAURENT, P. *Rev. de Mét.* 42 (1945) 79, 125, 156, 194, 230

[50] BRAGG, W. L. *Nature, Lond.* 149 (1942) 511
[51] WOOD, W. A. *Proc. roy. Soc. A* 172 (1939) 231
[52] —— *Proc. phys. Soc. Lond.* 52 (1940) 110
[53] SMITH, C. L. *Nature, Lond.* 160 (1947) 466; see also COTTRELL, A. H. and GIBBONS, D. F. *Ibid* 162 (1948) 488
[54] ZENER, C. *Met. Tech.* 13, *Tech. Pub. No.* 192 (1946)
[55] READ, T. A. *Phys. Rev.* 58 (1940) 371; *J. Applied Phys.* 12 (1941) 100; *Trans. Amer. Inst. min. met. Engrs* 143 (1941) 30
[56] ROSENHAIN, W. *J. Iron Steel Inst.* 70 (1906) 189
[57] READ, T. A. and TYNDALL, E. P. T. *J. Applied Phys.* 17 (1946) 713
[58] FETZ, E. *Trans. Amer. Soc. Met.* 25 (1937) 1030; 26 (1938) 961
[59] MILLER, J. V., BANNISTER, L. C. and HINDE, R. M. *Nature, Lond.* 158 (1946) 705
[60] BURGERS, W. G. *Proc. roy. Acad. Sci. Amsterdam* 50 (1947) 452, 595, 719, 858
[61] SANDEE, J., MAY, W. and BURGERS, W. G. Summarized in *Nature, Lond.* 157 (1946) 76
[62] ANDERSON, W. A. and MEHL, R. F. *Amer. Inst. min. Engrs* Tech. Pub. No. 1805 (1945)
[63] BURGERS, W. G. *Nature, Lond.* 160 (1947) 398
[64] GUINIER, A. *J. phys. Radium* 3 (1942) 124
[65] MOTT, N. F. and NABARRO, F. R. N. *Proc. phys. Soc. Lond.* 52 (1940) 86
[66] —— *J. Inst. Met.* 72 (1946) 367
[67] NABARRO, F. R. N. *Proc. phys. Soc. Lond.* 58 (1946) 669
[68] OROWAN, E. *Discussion. Symposium on Internal Stresses* Institute of Metals, London, 1947 p 451
[69] ANDRADE, E. N. DA C. *Proc. roy. Soc. A* 84 (1910) 1; *A* 90 (1914) 392
[70] COTTRELL, A. H. and AYTEKIN, V. *Nature, Lond.* 160 (1947) 328
[71] OROWAN, E. *J. West Scotland Iron and Steel Inst.* (1947)
[72] HANSON, D. and WHEELER, M. A. *J. Inst. Met.* 45 (1931) 229
[73] KAUZMANN, W. *Trans. Amer. Inst. min. met. Engrs* 143 (1941) 57
[74] SMITH, C. L. *Proc. phys. Soc. Lond.* 61 (1948) 201
[75] ANDRADE, E. N. DA C. *Science Progress* 30 (1936) 593
[76] OROWAN, E. *Int. Conf. Phys. II The Solid State of Matter* Physical Soc. London, 1934 p 81

3

CRYSTAL BOUNDARIES

R. King and Bruce Chalmers

KNOWLEDGE about the regions separating or joining the crystals of metals has until recently been obtained mainly as a by-product of experiments made for quite different reasons. There are still many aspects of the subject that are not understood, but recent work, both experimental and theoretical, has led to a clearer understanding of the part which the crystal boundaries play in the resistance of metals to deformation.

The purpose of this chapter is to review the most recent experimental and theoretical work on the behaviour of crystal boundaries. It is necessary to examine the new results in terms of the ideas and experiments of the past in order to see whether the older facts as well as the newer ones are accounted for.

Attention is focused in turn on the following aspects of the subject. The first problem is the origin of the boundaries themselves, after which the mechanical properties and effects of the boundaries are discussed, and a short account is given of the work which has been done on the melting of crystal boundaries. Crystal boundaries may for many purposes be regarded as surfaces, and they possess surface energy which may manifest itself as a surface tension. After considering the consequences of this, the chemical behaviour of the boundaries is described. The subject of intercrystalline corrosion has been excluded, as it throws little if any light on the main problem, the structure of the boundary. A final section summarizes the conclusions which may be drawn from the available facts.

THE ORIGIN OF BOUNDARIES

When a liquid metal is cooled, a temperature is reached at which the solid and liquid are in equilibrium; this is the freezing point. At lower temperatures the solid form is the more stable and there is a tendency for nuclei of the crystalline solid to form. As soon as nuclei are formed they increase in size until the temperature has risen, by virtue of the release of latent heat, to the melting point. Subsequent growth of the

crystals can only take place if heat is abstracted from them. In relation to crystal boundaries two cases may arise, *a* that in which the degree of supercooling is such that the whole of the liquid freezes before its temperature has risen to the melting point, *b* that in which the crystals do not come into contact with each other until heat is being extracted from the system. The two sets of conditions may occur in sequence in neighbouring points of the same material. This happens when the outer layer of an ingot freezes with an equiaxial structure (case *a*) and the remainder with a columnar structure (case *b*). The equiaxial zone that may form in the part of an ingot which freezes last is also due to the formation of nuclei in a supercooled liquid. The reason for the supercooling is different from, but the essential features in relation to the crystal boundaries are similar to, those of the chilled zone. The essential difference is that in case *b* an isothermal surface separating solid and liquid is geometrically determined by the loss of heat from the solidifying mass, while in case *a* external loss of heat is not involved.

It is necessary to consider whether there is any difference in the detailed mechanism of the formation of the crystal boundaries in these stages of solidification.

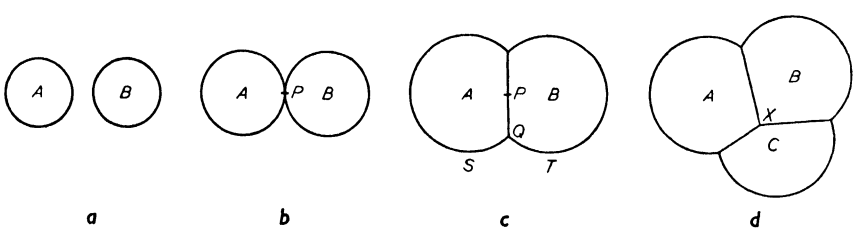


Figure 1. Diagrammatic representation of the successive stages in the formation of boundaries

In the first equiaxial stage each crystal will grow extremely rapidly and, in the first instance, according to its natural habit. As soon as neighbouring crystals begin to impinge upon each other a boundary will be formed. The situation is illustrated in *Figure 1a*, in which two crystals, A and B, are represented as growing spherically. It is possible that when crystals are growing rapidly as a result of the nucleation of a supercooled liquid they will grow spherically. In *Figure 1b* the crystals have just touched at the point P. *Figure 1c* shows a subsequent state of affairs in which the boundary PQ has been formed coincidentally with

the growth of the crystals. The surface represented by *SQT* is the instantaneous liquid/solid interface and is an isothermal surface at the melting point. It is interesting to note that here the liquid is cooler than the solid. The process continues until the boundary *PQ* meets the boundaries between the crystals *A* and *B* and a third crystal *C* (*Figure 1d*) at the point *X*. This stage of solidification can be reached at a very high rate since it is not dependent on the extraction of heat from the system. It is probable that there is not sufficient time available for liquid to move into the spaces left by the shrinkage which in most metals accompanies freezing. It is to be expected, therefore, that the boundaries will be 'short of atoms' and relatively weak.

In the second stage, that of columnar growth, the boundary is formed between two crystals which are both subject to heat loss in a definite direction. The surface of separation between solid and liquid advances in a regular manner and the boundary is formed simultaneously in such a way that there is always a supply of liquid at the point where the boundary is being formed. There should then be no lack of atoms to fill all the available sites in the boundary. Therefore, it may be concluded that there is probably a difference between the 'equiaxial' and the 'columnar' types of boundary. It is of interest to note that no differences in the properties of the two types of boundaries have been detected; the reason may be that experiments on the specific properties of boundaries have been confined largely to boundaries of the 'columnar' type.

The crystal boundaries found in ordinary samples of metals may differ in two important respects from those we have been considering. In the first place the metals of practical interest are nearly all alloys, in which more than one type of atom is involved, and secondly any metal that has been deformed in the process of fabrication is likely to have undergone some degree of recrystallization or grain growth. The boundaries, in these circumstances, are not in the positions in which they were formed during solidification, and their properties may have been modified by their movement or by their destruction and re-creation.

Recent experimental work[1] has shown that the mode of formation of the columnar type of boundary is influenced by the relative orientations of the two crystals concerned, and in particular by the relative orientations of the crystal axes to the surface of separation of solid and liquid. It has been found that for tin, lead, and zinc the boundary is formed

in a direction perpendicular to the interface only when the crystals have similar orientations *i.e.* when the crystallographic indices of the crystal surface in contact with the liquid are the same for the two crystals.

When this condition is not satisfied, the direction of formation of the boundary is oblique. In extreme cases the angle may be as high as 45°. The direction of the boundary is such that the crystal which has the closer atomic packing on the relevant surface encroaches on the other crystal. In order to account for this phenomenon it was necessary to postulate that the melting point varied with the nature of the face in contact with the liquid.

MECHANICAL PROPERTIES AND EFFECTS

It is perhaps in relation to their effects upon the mechanical behaviour of polycrystalline aggregates that grain boundaries have received most attention, and it is of interest to observe how theories concerning their nature and behaviour have grown in relation to knowledge of the mechanism of deformation within the crystals themselves.

The first theory of the boundary which was developed in detail and which with minor modifications was widely accepted for many years was the amorphous cement theory, of which ROSENHAIN was the chief protagonist. The presence of a cement at grain boundaries had been suggested by earlier writers. QUINCKE[2] and BRILLOUIN[3] both suggested a form of boundary cement, and EWING and ROSENHAIN[4] in 1901 had referred to 'the fusible and mobile eutectic forming the intercrystalline cement' even in pure metals. However, the amorphous cement theory took shape after the work of BEILBY. Beilby's[5] observations on the flow of metals in the polishing of surfaces led him to suggest that during plastic deformation localized melting and flow followed by solidification occurred within the crystals. The melted material solidified in an amorphous condition (it was a supercooled liquid) and had the characteristics of a vitreous material, and the change in properties caused by plastic deformation could be attributed to the presence of this amorphous material. BENGOUGH[6] first identified the grain boundary cement with Beilby's amorphous metal. He had carried out a series of tensile tests on a number of metals over a range of temperatures and observed that at a certain temperature characteristic of the metal there was an abrupt change in the curve relating strength to temperature. To explain the generally observed fact that, at room

temperature, fracture in metals and alloys was transcrystalline, which suggests that the cohesion between separate crystals was greater than that between different parts of the same crystal, he suggested that the boundary material 'must surely be none other than Beilby's amorphous material arranged in a thin, more or less continuous, layer round the crystals'. He postulated that this layer existed in all metals and alloys at ordinary temperatures, but above a certain temperature, at which an abrupt change occurred in his experimental curves, it could not exist even momentarily and fracture occurred between the crystalline junctions. Rosenhain,[7] while accepting the existence of a cementing film at room temperature, did not agree that the amorphous material ceased to exist above a certain temperature. He suggested that Bengough's results could be explained by the behaviour of the grain boundary material as a vitreous substance becoming a viscous liquid at high temperatures. In a later paper by ROSENHAIN and EWEN[8] the mechanical behaviour was elaborated as follows. 'An undercooled liquid while very hard and brittle at the ordinary temperature becomes soft and viscous at high temperatures, while the reduction of internal friction which opposes slip within the crystals does not diminish with such extreme rapidity with rising temperature. Up to a certain temperature we should thus expect the course of the temperature/strength curve to be governed primarily by the diminution of internal friction, but at that point the softening of the cementing liquid would overtake the softening of the crystals and from that temperature upwards the temperature/strength curve would follow the decrease of the viscosity of the liquid with increasing temperature.'

Following these early papers a number of observations were explained in terms of an amorphous layer at the boundary. In the section on 'Metals: the relations of strain and structure—amorphous metal' in the 1923 *Dictionary of Applied Physics* Rosenhain puts the case for the existence of an amorphous cement. It is very clear how much the amorphous cement theory of the boundaries was dependent upon the success of the Beilby hypothesis of plastic deformation in the crystals, and much of Rosenhain's argument consists of the evidence for the existence of amorphous material from these considerations.

Having established the plausibility of the existence of metal in an amorphous condition Rosenhain's argument for its existence at the grain boundaries was that its presence could explain the following

experimental facts. Bengough and subsequent workers observed that, for rapid deformation and at ordinary temperatures, fractures in most metals were transcrystalline except in cases where intercrystalline weakness arose, such as in the presence of a weak constituent. For slow deformation at higher temperatures, however, the fracture path tended to be intercrystalline. In almost all metals at ordinary temperatures superior mechanical properties were associated with a fine grained structure. Microscopic examination of the slip bands in the neighbourhood of grain boundaries showed that the bands were distorted in a manner that indicated that the crystal boundary interfered with the slip within the crystal. Observations on iron by ROSENHAIN and HUMFREY[9] showed that as the temperature of deformation approached 1,000°C the crystals showed less deformation by slip and there was more deformation at the boundary. Rosenhain and Ewen[10] showed that under slight stresses a number of metals failed at the boundaries a few degrees below the melting point, the crystals being pulled away without appreciable deformation. All these observations could be explained by the existence at the boundary of a material which was strong, hard, and quasi-elastic at ordinary temperatures but capable of softening relatively rapidly with increasing temperature and at higher temperatures plastically deformed by very small stresses. Such properties were known to characterize vitreous materials and had been ascribed to Beilby's amorphous metal, so the existence of a layer of such material at the boundary was a reasonable conclusion.

'Season cracking' was explained as the result of the slow viscous deformation of such a boundary material over long periods under the action of internal stresses, and evidence suggesting that it could occur more readily where boundaries were smooth was adduced in favour of this argument by ROSENHAIN and ARCHBUTT,[11] though as the smooth boundaries were produced by a particular form of heat treatment it could be readily argued that the susceptibility to season cracking might be due to a change induced by the heat treatment rather than the mere formation of straight boundaries.

Other chemical and physico-chemical evidence was put forward in support of the amorphous cement theory of the boundary, but this is referred to in other parts of this paper. Strong support was given by JEFFRIES,[12] who collected a considerable amount of information concerning the relative properties of amorphous and crystalline phases.

He concluded that the cohesion of an amorphous substance was substantially zero at the melting point, though the cohesion of the crystalline phase could be considerable. With decreasing temperature, however, the cohesion of the amorphous phase increased more rapidly than that of the crystalline phase. At a certain temperature (usually about 0·35 to 0·45 of the absolute melting point) the cohesion/temperature curves intersected. This he called the equicohesive temperature. He noted too that this equicohesive temperature corresponded closely to the lowest recrystallization temperature and argued that this was consistent with the concept of amorphous material at the boundary.

As a convenient measure of this equicohesive temperature Jeffries determined the temperature at which specimens with large grain size had the same strength as those of small grain size. In accordance with the supposition that the grain boundary was amorphous he found that the experimentally determined value of the equicohesive temperature was dependent upon the rate of straining: the slower the rate of straining, the lower the equicohesive temperature.

The concept of the amorphous cement gradually changed. For example, Rosenhain, who originally pictured it specifically as a supercooled liquid, later spoke[13] of an 'irregular intercrystal layer' rather than an amorphous cement and suggested that the boundary layer might not be wholly amorphous but might for instance consist of an amorphous matrix in which were dispersed minute crystallites. A similar type of boundary layer was suggested by GOUGH.[14] From his x-ray observations on fatigue he concluded that small but perfect crystallites were produced at the slip planes, possibly dispersed in an amorphous phase. These crystallites differed in orientation but not in lattice parameter, and apparently provided a means of transition between the two undeformed layers on opposite sides of the slip band. He pointed out the possibility that such a system of crystallites could form a means of transition from one lattice to another at the boundary between two crystals.

Rosenhain[13] suggested in 1927 that the only two boundary structures possible were a smooth curving transition from one orientation to another and a boundary in which an intermediate layer served to connect the differing orientations. He dismissed the idea of curving transition because, assuming that the allowable increase in distance

between atoms was less than the elastic limit range, the radius of curvature would be too large and the boundary in small crystals would occupy the whole crystal, whereas for crystals of 10^{-4} cm x-rays had shown a constant parameter over most of the crystal. On the other hand, the amorphous layer he estimated at 10 to 100 atoms thick. Nevertheless, in 1929 HARGREAVES and HILLS[15] proposed a transition lattice for the boundary. They argued that in an aggregate of crystals the forces which caused the atoms to assume the regular arrangement of the crystal lattice must still be operative at the boundaries, causing the atoms at the boundaries to take up positions dictated by forces exerted by the crystals on either side. Thus in unstrained metals there would be a definite pattern of atoms which was always the same for the same two relative orientations and the same relative position of the boundary. In considering the behaviour of such a boundary under stress, they suggested that since it possessed a definite structure it should possess an elastic limit and plastic flow should not take place until the limit had been exceeded. Plastic flow would result in the middle of the boundary having an amorphous structure which could be considered to possess the properties associated with viscous undercooled liquids. On each side of this amorphous layer there would be a zone of transition to the crystalline lattice. The greater the amount of mechanical working, the thicker the amorphous layer. The explanation of the mechanical behaviour of metals then followed as for the amorphous cement theory. As evidence for the widening of the boundary Hargreaves and Hills cited the work of TOWNSEND[16] who found that fatigue caused an apparent widening of the crystal boundaries in an alloy of lead containing 1 per cent antimony. The widening had been ascribed to the reduction in solubility of the antimony, but Hargreaves and Hills pointed out that whatever the detailed explanation of the fact it was evidence that some change took place which extended into the crystals on each side of the boundary.

Annealing, it was suggested, resulted in either the disappearance of the amorphous layer followed by grain growth as the transition zone swept forward into the more distorted crystal or, with greater deformation and consequently wider amorphous layer, the formation of new crystals at the boundaries.

Rosenhain considered that the arrangement of the atoms suggested by Hargreaves and Hills in the unstrained metal did not fulfil the conditions

necessary for stable equilibrium. In some cases the atoms would be brought too close together and in others they would be too far apart to fulfil the conditions of atomic linkage existing in metals. It may be noted, however, that in describing the amorphous cement Rosenhain himself had said that 'in the irregularly arranged layers the atoms will be found to lie at all manner of distances ranging from the minimum to the maximum across which interatomic cohesion can occur, while there will be many instances where adjacent atoms are too far apart to be physically connected at all'.

Since those days the accumulation of knowledge, particularly from x-ray investigations, concerning the structure and behaviour of metallic crystals has made less plausible the concept of a separate amorphous phase at the grain boundaries. The processes of deformation have been studied extensively and there is no longer any need to postulate the formation of a Beilby amorphous layer to account for the changes in hardness which accompany deformation. On purely theoretical grounds the existence of a stable amorphous state over relatively large volumes in metals is extremely unlikely. BRAGG[17] puts forward a strong argument against its existence. He points out that interdiffusion takes place quite rapidly in metals at relatively low temperatures, and such diffusion implies the movement of atoms over relatively large distances. On the other hand, the movements required to transform the amorphous arrangement of metal atoms into a regular crystalline pattern are very small and may be expected to occur at temperatures considerably lower than those at which diffusion is appreciable. Bragg states that 'a macroscopic Beilby layer between the crystallites is very difficult to conceive, for the atoms in the centre of it, being quite free to move, would immediately take up a regular crystalline form. An amorphous solid such as vitreous silica can be frozen into a state of almost indefinite permanence because the bonds between neighbouring atoms are so strong. Once formed in an irregular way as the liquid cools they become permanent at lower temperatures, since the chance of a thermal fluctuation being sufficiently powerful to break all the bonds and permit rearrangement in any localized area is vanishingly small. On the other hand, in a metal kT is sufficiently great to permit rapid rearrangement right down to room temperature.' Here Bragg is referring to cold worked material in which each grain is pictured as consisting of a mass of differently orientated units or

crystallites of dimensions of the order 10^{-4} cm. The argument applies equally to the case of the boundary between grains.

The study of solid solution and precipitation phenomena has provided a considerable body of information concerning the elastic strains which are possible within a crystalline lattice. The rapid and successful development of the theory of dislocations[18] has focused attention on the distortions which can occur, and it is now clear that the strains which would be associated with a comparatively rapid transition from one crystal orientation to another at a grain boundary in a form very similar to that suggested by Hargreaves and Hills would be in no way incompatible with the properties of metal crystals. Thus a form of transition layer is now widely accepted as the most likely boundary structure on theoretical grounds.

BURGERS,[19] in discussing the boundaries between 'mosaic blocks' in an imperfect crystal, describes how the transition of orientation from one crystal block to another may be brought about by a system of dislocation lines. The simple case of a cubic lattice is treated, illustrated by reference to two crystal blocks which are rotated relative to one another about one cubic axis. The dislocations are regularly spaced with a period h, and for small angles the angle between the two crystal blocks is λ/h, where λ is the interplanar spacing indicated (see Chapter 2, *Figure 10*). For crystalline blocks rotated relative to one another about the normal to the boundary between them the transition from one orientation to the other may be considered as a system of Burgers, screw type, dislocations. The examples considered are the simplest forms of boundary, but formally any boundary between two crystal blocks may be built up from a combination of these two elementary types of transition. While the treatment given by Burgers applies to the transition regions between blocks within a single crystal, the picture is equally applicable to intercrystalline boundaries. While such a formal representation of the boundary contributes little towards the understanding of its mechanical behaviour, though it offers a means of estimating the free energy of the boundary in simple cases, it clearly indicates the plausibility of a transition picture of the boundary, and it is of interest to observe how the picture of the grain boundary has been associated with the theories concerning the methods of deformation of the crystals themselves. When the Beilby hypothesis of work hardening was popular it led to the hypothesis of the amorphous cement. The

dislocation hypothesis of deformation has led to the formal representation of the grain boundary as a system of dislocations.

Before we consider how the transition concept of the grain boundary, theoretically the most probable form, may be related to experimental facts it is useful to compare the implications of the two pictures. In its earliest form the amorphous cement theory postulated the existence of a distinct phase at the grain boundary. This phase was of the nature of a supercooled liquid, although later it came to be less closely defined as an irregular arrangement of atoms. Whatever the detailed picture of the atomic arrangement, the positions of the atoms were regarded as random, the boundary phase having the properties of a vitreous material. In the transition picture the atoms are again regarded as irregularly arranged in the sense that they do not lie on crystal lattice points, but there exists a regularity in that their positions are dictated by the demands of the two crystals meeting at the boundary, and the atomic arrangement will be the same for any two boundaries where the orientations of the crystals and the direction of the boundary are the same. On the transition hypothesis the degree of departure from lattice positions will depend upon the difficulty of bridging the change of orientations between the grains, and it would be anticipated that the properties of a boundary would depend upon the relative orientations of the grains meeting at the boundary and on the direction of the boundary itself. If on the other hand there existed a truly amorphous phase at the boundary, the properties of the boundary should not vary with the orientations of the grains. Experiments designed to determine differences in boundary properties with the orientations of the crystals should give evidence concerning the exact nature of the boundary, but so far such experiments have been somewhat inconclusive and incomplete.

In discussing experimental observations on the mechanical behaviour of grain boundaries it is necessary to distinguish between actual properties of grain boundaries and their effects, that is the interaction of neighbouring grains across the boundary arising from the proximity of two grains. These effects are clearly dependent on the nature of the grain boundary material. It is impossible to transmit a stress of any nature from one grain to another unless the boundary itself is capable of withstanding a stress of that nature, but the result is not a measure of the boundary properties themselves. Because polycrystalline

materials in general show a higher yield strength than single crystals it does not follow that there exists at the crystal boundaries a material of higher yield strength than the crystalline materials themselves. All that may be concluded is that the grains in a polycrystalline aggregate cannot deform plastically as they would if they were isolated, because if they did they would no longer conform with each other at the boundary. The grains thus interfere with each other across the boundary.

The inference which may be drawn concerning the boundary itself is thus limited, we can conclude only that it is such as to allow this mutual interference to take place.

The mutual interaction of grains has been observed by a number of workers. CARPENTER and ELAM[20] observed the supporting effect of grains upon each other in the tensile deformation of coarse grained aluminium, the grains being less deformed near the boundary than elsewhere. This effect extended some distance into the crystal grains and was observed as far as 0·25 cm from the boundary. ASTON[21] found that the crystal grains suffered less change of orientation at grain boundaries than within the grains themselves and concluded that less deformation occurred in the boundary regions. BARRETT and LEVENSON[22] investigated the behaviour of aluminium in compression and showed that the deformation within grains was far from homogeneous and depended upon the surrounding grains.

As an example of a straightforward investigation of such effects of the grain boundary the experiments of CHALMERS[23] may be quoted. The variation of yield stress in tension of cylindrical specimens of tin containing two crystals was examined. The boundary between the two crystals was longitudinal and contained the axis of the specimen. The specimens were grown so that in any given specimen one crystal was the mirror image in the boundary of the other. In all specimens the crystals had the same orientation relative to the axis of the specimen, but the angle between the two crystals varied from specimen to specimen. Thus when the specimen was stressed longitudinally the stress was in the same direction relative to the crystal axes of the grains, and the only difference from specimen to specimen was in the angle between the grains. The stress required to produce a definite small increase in length was determined for a number of specimens covering a wide range of angles between the crystals. It was found that this

stress varied approximately linearly with the angle between the crystals, and extrapolation to zero angle gave the value found for a single crystal of the same orientation relative to the stress. The stress varied from 400 gm/mm^2 for the single crystal to 650 gm/mm^2 where the angle between the grains was 90°. This change in yield stress cannot be attributed to the grain boundary itself. It arises from mutual interference of the grains which is thus seen to be considerable.

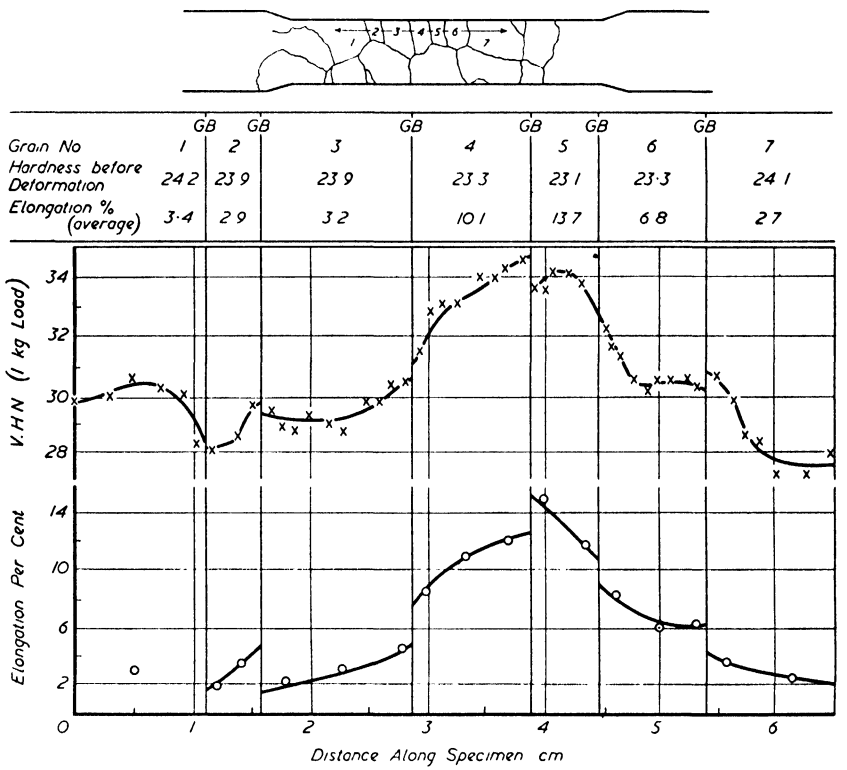


Figure 2. Hardness and local elongation of the grains of an aggregate after five per cent elongation. Slip occurring at various points in the grain is shown schematically

More recent experiments of BOAS and HARGREAVES[24] demonstrate in a convincing manner the interaction of grains in a polycrystalline aggregate. Specimens of aluminium with a grain size of 1 to 3 cm were electrolytically polished and lightly etched. Vickers Diamond Hardness measurements were made in a straight line traverse across the grains and the specimens deformed in tension. The elongation of different

portions of each grain was determined from the change in distance between the indentations, and the hardness at different points after deformation was determined by a second traverse parallel to the first. It was found that the elongation was different in different grains and that deformation was not uniform within any one grain. *Figure 2* is typical of their results. Whereas before deformation the hardness was constant within any one grain, it was found to vary considerably after deformation. There was an obvious tendency to conform in elongation at the grain boundary. A grain which on the average suffered little deformation tended to show increased deformation in the neighbourhood of its boundary with a grain which suffered greater average deformation, and *vice versa*. The hardness varied similarly and supported this observation. Observations on slip lines and x-ray examination confirmed the nature of the inhomogeneity of deformation. Following Chalmers,[25] Boas and Hargreaves conclude that owing to this mutual interference of neighbouring grains, regions of the grains near the boundary must deform by a process which is not typical of the deformation in the centre of the grain. TAYLOR[26] considered the deformation of a polycrystalline material on the assumption that the mutual interference is such that each grain is constrained to deform in the same manner as the aggregate as a whole, and showed that any arbitrary deformation would require slip in five different slip systems. There was, however, no evidence of fivefold slip in Boas and Hargreaves experiments. They argue that if, as appeared, the deformations tend to be continuous at the grain boundary, deformation by slip would mean different stresses on opposite sides of the boundary. Since the stresses must be continuous across the boundary, there must exist a mode of deformation which is not ordinary slip.

The mutual interference of grains has been observed at elevated temperatures and slow rates of straining. HANSON and WHEELER[27] stressed in tension at 250°C a specimen of aluminium consisting of two crystals. The boundary between the crystals was roughly perpendicular to the direction of the stress, and while normal slip bands appeared in both crystals, in one of the crystals there appeared wavy distortion marks near the boundary and signs of slip in planes different from those operating away from the boundary. The cross-section of the specimen did not decrease at the boundary to the same extent as in the crystal grains, and the specimen assumed a barrelled appearance in this region.

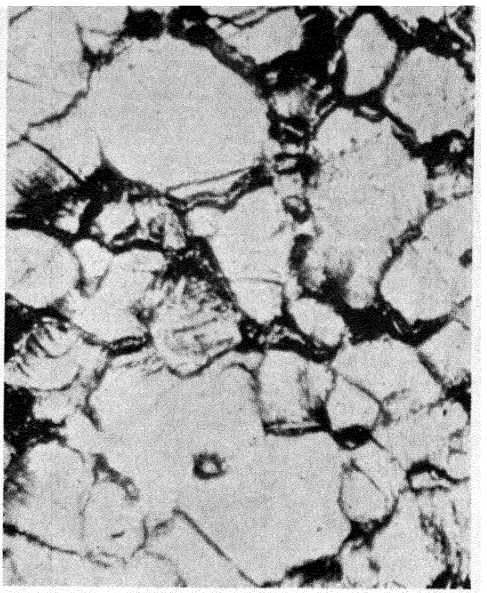


Plate IV. Aluminium specimen after 210 hours under a stress of 1·4 tons/sq in at 250°C. Elongation 21 per cent (× 150)



Plate V. Specimen of 2 per cent lead-tin alloy after 1,110 hours under a stress of 1,000 lb/sq in. Elongation $3\frac{2}{3}$ per cent (× 400)

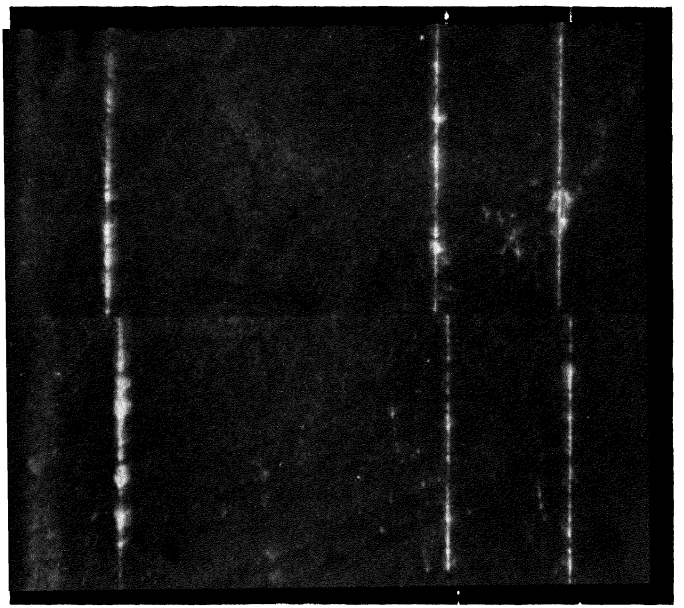


Plate VI. Shear movement at the grain boundary in a tin bi-crystal after 50 hours under a shear stress of 590 gm/sq cm at 222°C (× 20)

These then are typical effects of grain boundaries. Their importance is obvious, but it is necessary to distinguish clearly between these and the actual properties of the boundary material.

Although the transition picture of the boundary is the more attractive on theoretical grounds, it is necessary to see how far it is possible to relate it to experimental observations. It was seen that the amorphous cement hypothesis fitted well qualitatively with the observed facts. The facts which have to be explained are very much those which Rosenhain quoted in support of his hypothesis in 1923. At ordinary temperatures and speeds of straining below a certain temperature characteristic of the metal, fracture is transcrystalline. At slower rates and above that temperature there is deformation at the boundaries and failure tends to be intercrystalline.

The technological importance of creep phenomena has led to many investigations of the behaviour of metals at slow rates of deformation at elevated temperatures, and a considerable amount of attention has been paid to the role played by the grain boundary in the deformation. The importance of the grain boundary has been stressed by many observers, as for example HANSON and WHEELER,[27] MCKEOWN,[28] GREENWOOD and WORNER,[29] and some of the results of Hanson and Wheeler may be quoted as typical. Specimens of aluminium were polished and examined microscopically after various types of deformation. In an annealed specimen of uniform grain size, stressed at 1·4 tons/sq in at 250°C, after the initial rapid straining on application of the load, when there was some slip in the grains, extension was very slow and no marked changes occurred in the microstructure. Later, however, the rate of extension increased, and after 210 hours the grain boundaries appeared to be broadened, and there was much movement in their neighbourhood (*Plate IV*). A specimen strained rapidly at the same temperature to the same elongation showed only slight indications at the boundaries but very much slip within the grains. When the specimen under low stress was loaded for a long time, cracks appeared between the grains both on the surface and within the specimen, and gradually extended with time until fracture occurred. The fracture, as would be expected, was not entirely intercrystalline. As the intercrystalline cracks developed the effective cross-sectional area would be reduced and the stress would increase until it became large enough to cause rapid fracture, which would then be transcrystalline.

however, it is necessary to consider only very small displacements as it is clear that, where a boundary is not straight or where there are interlocking grains, after a very small shear movement at the boundary further movement will be blocked by mutual interference of the grains.

Internal friction is a case in which deformations may be very small, and the investigation of the variation of internal friction with temperature and grain size by BARNES and ZENER[33] and ZENER, VAN WINKLE, and NIELSEN[34] has yielded results consistent with the hypothesis that boundaries behave in a viscous manner and allow the relaxation of shear stress across them. The most complete series of experiments designed to investigate the viscous behaviour of boundaries when deformed by small amounts are those of Kê.[35] He measured a number of so-called anelastic effects in wires of aluminium of 99·991 per cent purity. From the fact that the observed effects in polycrystalline specimens were linear with respect to applied stress and prior strain and could be correlated by the application of the Boltzmann superposition principle he concluded that they could be explained on the assumption that the material consisted of two components, one truly elastic and the other behaving in a viscous manner. As the effects attributable to the viscous component were absent in single crystal specimens he concluded that the viscous deformation in the metal occurred at the grain boundaries. In the first instance he investigated four anelastic effects in torsion:

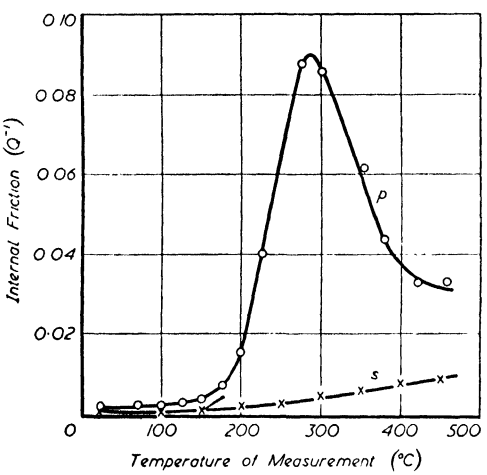


Figure 4. Variation of internal friction with temperature in polycrystalline (p) *and 'single crystal'* (s) *aluminium. Frequency of vibration = 0·8 cycle per second at room temperature*

a variation of internal friction with temperature
b variation of dynamic rigidity with temperature
c creep under constant stress
d relaxation of stress at constant strain.

These phenomena would clearly be affected by viscous behaviour

at the grain boundaries. For the investigations it was necessary that stresses should be low enough for all effects to be entirely recoverable after the removal of the stress, and that internal friction and dynamic rigidity should be independent of amplitude of vibration. Thus the deformations were limited to strains of the order 10^{-5}.

With a frequency of vibration of 0·8 cycle per second the internal friction varied with temperature as indicated in *Figure 4*. On the assumption of viscous shear at the grain boundary the existence of a maximum was to be expected. Since the internal friction is determined by the product of the distance moved and the shear stress necessary to cause the movement, it would be small at low temperatures where the movement would be small and at high temperatures where the resistance to slip was small. At an intermediate temperature, however, where both slip and resistance to slip were not small, the internal friction would be expected to be appreciable. The effect was observed in polycrystalline specimens but not in single crystals.

The variation of rigidity with temperature is indicated in *Figure 4*. The figure relates the square of the frequency, which is approximately proportional to the rigidity, and the temperature. Compared with the single crystal specimen the rigidity of the polycrystalline specimen shows a marked drop above a certain temperature. Such a drop would be expected if the boundary behaved in a viscous manner. As the temperature was raised, so would the boundary become more mobile until at a certain temperature appreciable boundary slip would occur during the period of an oscillation. Kê found that the value of the fully relaxed dynamic rigidity obtained from his experiments was in close agreement with that calculated in a manner similar to that of ZENER[36] for the value of Young's modulus, on the assumption that boundaries were incapable of supporting a shear stress.

Torsional creep under constant stress up to a maximum surface strain of 2×10^{-5} was observed at different temperatures. The creep was proportional to the stress and was completely recoverable after the removal of the stress. No such effects were observed in single crystals. Kê showed that the creep could be expressed as

$$d_t/d_o = A \,.\, \mathrm{f}(te^{-H/RT}) \qquad \ldots\ldots (1)$$

where d_t was the deflection after time t, d_o the deflection immediately after the application of the load, and A a constant independent of the

time or the temperature T. The value of the activation energy H was determined as 34,000 calorie/gm atom.

The relaxation of stress for a constant strain was investigated at different temperatures and it was found that an activation energy for the process could be determined. This was 34,500 calorie/gm atom, in close agreement with the value calculated from the creep. The total amount of stress relaxation produced was in accordance with the observed reduction in dynamic rigidity with temperature.

Using the Boltzmann superposition principle, Zener has derived relationships between the stress relaxation at constant strain and each of the other effects: dynamic rigidity, internal friction, and creep under constant stress. Using these relations Kê calculated values for the stress relaxation for different times at 200°C, in turn from his observed values of dynamic rigidity, internal friction, and creep. The results calculated from each type of observation were all in close agreement with the experimentally determined stress relaxation. Thus all the observations gave a self consistent picture indicating that the grain boundaries behaved in a viscous manner. On this assumption Kê made an estimate of the coefficient of viscosity of the boundary. The temperature dependence of viscosity was given directly by the temperature dependence of the creep process, and the actual value at one temperature (285°C) could be estimated from the maximum of the internal friction temperature curve. The value estimated for the coefficient will, of course, depend upon the effective thickness of the boundary, and this is not known. Kê assumed a value of one atomic distance, 4 Å. This assumption led to the expression for the coefficient of viscosity:

$$\eta = 1{\cdot}2 \times 10^{-20} G_u \, e^{(34,500/RT)} \qquad \ldots . \quad (2)$$

where G_u is the unrelaxed rigidity of the material. Extrapolation of this expression to 670°C gave a viscosity value of 0·14 poise, in fair agreement with the value determined experimentally by POLYAK and SERGEEV[37] for liquid aluminium.

Later experiments by Kê[35b] gave further strong support to the conclusion that grain boundaries behave in a viscous manner. Such behaviour should mean that if the frequency of vibration were increased, the maximum in the internal friction/temperature curve would be displaced towards a higher temperature, and similarly an increase in frequency should raise the temperature at which relaxation

of the rigidity occurred (*Figure 5*). This is clear from the following considerations. The internal friction due to relaxation across grain boundaries is appreciable when the period of oscillation is comparable with the relaxation time of the stress relaxation at the boundaries. If then the period of vibration is reduced the maximum internal friction will occur at a temperature where the relaxation time is smaller, that is, where the grain boundary is less viscous. This means a higher temperature. Similarly it requires a higher temperature for the grain boundary to deform sufficiently in the shorter period to completely relax the rigidity. Experimental observations at the different frequencies confirmed these expectations. Calculations of the activation energy for boundary slip from these observations gave in each case a value 32,000 calorie/gm atom, in good agreement with those obtained from creep and stress relaxation determinations. Observations on the effect of grain size are recorded and, it is claimed, are in good agreement with the postulate of viscous behaviour. Kê considers that the relaxation time associated with the stress relaxation across grain boundaries will increase with grain size because, for a given average stress, with the larger grain size, a larger relative displacement should occur before slip is prevented by interference of the grains at grain edges and corners, and thus a longer time will be required to reach the equilibrium state at which further grain boundary slip is completely blocked. Thus an increase in grain size should cause the internal friction/temperature curves and the rigidity/relaxation curves to be displaced in the direction of higher temperature. This argument is difficult to reconcile with observations carried out by King, Cahn, and Chalmers[31] on boundary shear in bicrystals of tin. Here, apparently, there are straight boundaries with no corners or grain edges to block the shear movement and yet such movement is limited, from which it is concluded that the limit

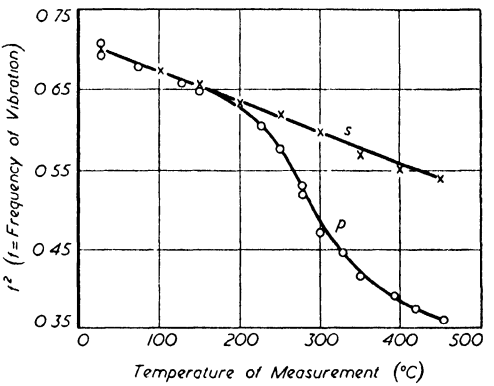


Figure 5. Variation of 'rigidity' with temperature, for polycrystalline (p) *and 'single crystal'* (s) *aluminium*

to shear movement is probably set by small irregularities in the boundary. Such irregularities should be present in the grain boundaries in polycrystalline specimens, and there is no reason to suppose that their spacing and hence the extent of movement is proportional to grain size. However, Kê's experimental results are in full accord with his argument and with an activation energy of 32,000 calorie/gm atom for grain boundary slip.

The observations of Kê have been considered in detail because they constitute the only quantitative results which have been concerned with the mechanical properties of the boundaries themselves as distinct from boundary effects, and as such represent an important advance.

At first sight Kê's results might be taken as strong support for the existence of an amorphous layer at the boundary. The fact that the extrapolated value of the viscosity a little above the melting point agrees with the experimentally determined value for liquid aluminium suggests the existence of supercooled liquid. However, the temperature dependence of the boundary viscosity is some 10 to 20 times greater than that of the liquid metal, so some other explanation must be sought. MOTT[38] has treated theoretically the behaviour of grain boundaries under stress and has suggested a mechanism whereby a region of transition from one crystal to another could behave in a manner consistent with Kê's observed results. He suggests that the boundary surface may be divided into regions or islands over which the fit between the two crystals is reasonably good, separated by lines where the fit is bad. The elementary process of slip is the relative movement of the two crystal surfaces constituting one of these islands of good fit. If the number of atoms constituting the island of good fit is n and the activation energy per atom for relative movement to occur is U, the activation energy for the elementary slip process is nU. In the presence of a shear stress σ the activation energy becomes $(nU - \frac{1}{2}\sigma la)$ for slip in the direction of the stress, and $(nU + \frac{1}{2}\sigma la)$ for slip in the opposite direction, a being the area of the island and l the distance moved in one elementary slip process. The probability that thermal fluctuations should supply an activation energy Q being $e^{-Q/kT}$, the probable number of times per second that slip occurs in the direction of stress is

$$\nu e^{-(nU - \frac{1}{2}\sigma la)/kT}$$

and in the reverse direction

$$\nu e^{-(nU + \frac{1}{2}\sigma la)/kT}$$

where ν is a constant of the order of magnitude of the frequency of atomic vibrations. For small stresses the net velocity of movement v in cm/sec in the direction of stress is shown to be:

$$v = (\nu l^2 a\sigma/kT)e^{-nU/kT} \quad \ldots\ldots (3)$$

This expression is in qualitative agreement with Kê's observations in which the velocity could be expressed as:

$$v = 18\sigma e^{-34{,}000/RT} \quad \ldots\ldots (4)$$

The activation energy nU may then be identified with the experimentally determined value. If U is estimated on the assumption that the two crystal surfaces forming the island simply slide over one another, the value of n giving agreement with the observed activation energy is approximately 23. This determines the value of a in the coefficient $\nu l^2 a/kT$, which is then $\simeq 6 \times 10^{-5}$. This value is very much smaller than the observed value 18(cm/sec)/(dyne/cm^2). Mott then suggests that the elementary process which enables slip to occur is the disordering of the atoms around the island. The free energy for this will vary with temperature, and since it must be zero at the melting point it is suggested that $U = L(1 - T/T_m)$ is a plausible relationship, L being the latent heat of fusion per atom and T_m the absolute melting point. Substitution for U in the expression for the velocity of movement gives:

$$\begin{aligned} v &= (\nu l^2 a\sigma/kT)e^{-nL(1-T/T_m)/kT} \\ &= (\nu l^2 a\sigma/kT)e^{nL/kT_m} e^{-nL/kT} \quad \ldots\ldots (5) \end{aligned}$$

The number of atoms to be disordered to give agreement with the observed activation energy is for aluminium about 14, and the coefficient $\nu l^2 a/kT$ is multiplied by a large factor e^{nL/kT_m}, which brings it more closely in line with the observed value.

Besides giving good quantitative agreement with the observed results, this picture has the important merit that it shows that the behaviour of the boundary approximates to that of a liquid layer at the melting point, because the activation energy for melting small volumes disappears at that temperature.

Thus the observed facts may be explained on the assumption that at the grain boundary the atoms fit both crystal lattices as best they can. There are regions where fit is good, separated by regions of bad fit. Thermal fluctuations cause disordering of atoms equivalent to local melting of small volumes, and in the absence of the applied stress

relative movement of the surfaces in contact occurs in random directions. The presence of an applied stress exerts a small directing influence on the relative motions which occur and the boundary deforms in a viscous manner. At temperatures which are relatively low compared with the melting point of the material the probability that such fluctuations will occur is extremely small and the boundary is very resistant to movement. When the temperature is sufficiently high for local disordering to occur to an appreciable extent, boundary slip will occur at a rate proportional to the stress and appreciable deformation of an aggregate will take place under small stresses, if sufficient time is allowed. There exists a limit to the amount of relative movement that can occur directly in this way. After a certain small movement, irregularities in the crystal surfaces, crystal corners, and edges will interlock and prevent any further simple slip. However, this interlocking will cause stress magnification at the irregularities and corners, and although the average stress applied to the specimen may be too small to cause plastic deformation within the crystals themselves, the magnification occurring at the irregularities will be large enough to deform the neighbouring crystalline material, giving an overall deformation due to the viscous behaviour of the boundaries. Such movements are no doubt responsible for the comparatively gross relative movement at the boundaries observed by Hanson and Wheeler and others. The manner in which this mechanism may contribute to the viscous component of creep and to subsequent tertiary creep is clearly apparent. It is not suggested that this is the only mechanism in these phenomena, but there can be no doubt that its contribution is considerable.

It is not difficult to see how, under these conditions of atomic rearrangement and stress concentration, fissures may form and spread at the grain boundaries, leading to the observed initiation of fracture by intercrystalline cracking.

MELTING OF CRYSTAL BOUNDARIES

One of the few specific properties of the crystal boundary that have been measured is the temperature at which it undergoes an abrupt change in mechanical properties which can be described as 'fusion'.

CHALMERS[39] prepared specimens of tin consisting of two crystals and one boundary. Since the two crystals were grown from separate 'seed' crystals it was possible to control the orientation of both of

them. It follows that the relationship between the crystallographic axes of one crystal and those of the other could also be controlled. The experiments consisted in subjecting such a specimen to a tensile stress across the boundary and raising the temperature slowly. The temperature at which the two crystals separated was measured, and it was observed that the difference between this temperature and the melting point of the crystals themselves had a constant value of 0·14C° over quite a wide range of conditions, independent of stress and rate of heating. It did not change when the angle between the *c* axes of the two crystals was varied between 14° and 85°. It had a constant value for tin of purities 99·998 per cent, 99·996 per cent and 99·986 per cent, while its value only changed to 0·24C° when an addition of 0·2 per cent of lead was made to tin of 99·986 per cent purity. The only type of boundary which did not exhibit the depressed melting point was that between twin crystals.

Fusion of the crystal boundaries at temperatures below the melting point of the crystals has also been reported by CHAUDRON, LACOMBE, and YANNAQUIS[40] for pure aluminium. The aluminium used in this work had purities described as 99·99 per cent and 99·998 per cent. It can be deduced from one of the illustrations to the paper that the depression of the melting point of the boundaries is about 0·25C°. From the photograph it appears that all the boundaries at the same temperature are equally affected. Further, it is stated that twin boundaries do not share this property with the boundaries between crystals of unrelated orientations.

The evidence available may be taken to show that there is a real difference between the melting point of the crystal boundary and that of the crystal, that this is not caused by impurities, and that it is independent of the relative orientation of the two crystals between which the boundary occurs.

All these experimental conclusions would be accounted for on the amorphous cement theory, but the constancy of the boundary melting point with varying relative orientation can also be explained on the transitional lattice theory if it is assumed that an atom in the transitional layer can have a maximum strain energy. Variation of the relative orientations would merely change the number of atoms with the maximum energy but would not affect the melting point, as this would be controlled by the maximum strain energy. The strain energy

corresponding to the observed depression of the melting point was found to be of the order 10^5 ergs/cm^3.

SURFACE EFFECTS

A considerable amount of information about crystal boundaries has been obtained from a study of the region where they meet the surface of the metal. The surface under discussion may have been formed during the freezing of the metal or it may have been formed after freezing, for example by polishing and etching.

It is necessary to distinguish between two kinds of information that can be obtained in this way. Firstly the examination of a surface at which crystal boundaries terminate may reveal inherent characteristics of particular boundaries or systems of boundaries, including their geometry both as they are formed and as they are affected by mechanical and thermal conditions. Secondly, evidence may be found of interactions occurring between a boundary and a free surface at which it terminates. Of these, only the second category can be described strictly as surface effects. The most important of these is the phenomenon of thermal etching. If a metal is heated to a suitable temperature, changes take place in its surface configuration. These changes include the appearance of a network of grooves, which coincide with the positions of the boundaries.

The phenomena were first reported by ROSENHAIN and HUMPHREY,[41] who observed the formation of two different networks on iron heated at a pressure equivalent to 10^{-2} mm of mercury. The two networks were later identified[42] as corresponding to the boundaries of the α iron and the γ iron crystals. The earliest serious investigation of the phenomena, however, was that of ROSENHAIN and EWEN,[8] who heated polished specimens of copper, silver, and zinc at a pressure of 2×10^{-3} mm and found that grooves developed at the grain boundaries but not at twin boundaries. Having demonstrated that more rapid loss of weight occurred in fine grained than in coarse grained material under the same conditions they concluded that the formation of the grooves was associated with evaporation. The following argument shows that the enhanced evaporation cannot be due to the excess vapour pressure of the assumed amorphous boundary material. In the first place, results of recent experiments on the relative movement under stress of neighbouring crystals[31] make it clear that the viscosity

to be attributed to the boundary material must be much higher than that derived by extrapolation of the viscosity of the liquid, and thus the amorphous boundary material would have a much higher viscosity than the supercooled liquid at the same temperature. The degree of disorder of the actual grain boundary material must therefore be less than that of the supercooled liquid, and it follows that the vapour pressure is also less than that which can be calculated for the supercooled liquid. The calculation has been made for silver, and it is found that the ratio of the vapour pressures of supercooled liquid to solid changes from 1·000 at the melting point, 961°C, to 1·042 at 900°C, 1·130 at 800°C, 1·250 at 700°C, and 1·420 at 600°C. Thus at temperatures which are high enough for appreciable evaporation to occur the vapour pressure of the boundary material cannot exceed that of the solid by more than an insignificant amount.

It was believed by Rosenhain and Ewen that the greater weight loss of the finer grain material was enhanced by the exposure of a greater area of metal due to the initial greater rate of evaporation at the grain boundaries. Thus the two processes would combine to produce the observed changes. It can be shown, however, that the formation of a groove does not increase the total rate of evaporation, as it is the projected rather than the actual area of the surface which should be taken into account, if the asperities are not considerably larger than the mean free path of the evaporating atoms.

It must be recognized, therefore, that the explanation advanced by Rosenhain and Ewen is not tenable, although the existence of the phenomenon is confirmed by the more recent work of FONDA[43] on tungsten. No acceptable hypothesis has been advanced, although the formation of boundary grooves has been observed and studied by various workers.

CARPENTER and ELAM,[44] for example, found the same sort of marking on an antimony-tin alloy, after heating to between 150°C and 200°C. From observation of the behaviour in relation to grain growth they concluded that the markings occurred during cooling. BOAS and HONEYCOMBE[45] account for their own similar observation, and for those of Carpenter and Elam, in terms of the stresses introduced by the differential expansion of neighbouring grains which have different thermal expansion coefficients parallel to their mutual boundary. Such effects, although of great interest, do not contribute to our knowledge of the crystal boundary, and cannot therefore be discussed here.

The formation of a boundary network in silver during exposure to elevated temperatures has been investigated by CHALMERS, KING, and SHUTTLEWORTH.[46] They showed that even at temperatures as low as 300°C a well marked but incomplete boundary network appeared. The pattern became complete at higher temperatures and after longer times. A careful analysis of the possible mechanisms led to the conclusion that the boundary grooves represent an approach to the equilibrium shape of the surface, which is depressed near the boundaries by the tension they exert. This is a manifestation of their surface energies. A knowledge of the angle at the base of a boundary groove would reveal the relative surface energies of the boundary and metal surfaces. Such measurements have not been made yet, but it is probable that the energies are of the same order of magnitude. It is believed that the shape of the surface approaches equilibrium by surface migration of atoms rather than by evaporation. A further deduction from observations on the thermal etching of silver is that twin boundaries have much less energy than ordinary intercrystalline boundaries, since the boundary grooves which are formed at the twin boundaries are much feebler, but nevertheless visible.

Information about the positions and movements of boundaries can be obtained from observations of the position of the intersections of the boundary with a particular plane, *i.e.* a surface which is prepared for examination. Thus it is observed that the traces of boundaries in such surfaces are often straight, or nearly straight, and it is concluded that the boundaries themselves are usually, although not necessarily, approximately plane. Further, it is usual for the traces of three boundaries to meet at a point; this suggests that each boundary is normally terminated by its junction with two other boundaries in a line. The angles between three such boundary surfaces should indicate the relative energies of the boundaries if equilibrium has been established. In order to ensure that the angles between the traces of the boundaries actually represent the angles between the boundary surfaces, it is necessary that the boundaries should be perpendicular to the surface on which the traces are examined. Experiments in which this condition was observed were made by CHALMERS.[1] The specimens of tin consisted of three crystals grown from seed crystals of the desired orientations which were so arranged that the three boundaries, at their junction with the surface, met at a point. This allowed considerable

freedom to vary the relative orientations of the crystals, and in particular it was possible to have a small difference of orientation, corresponding to a relative rotation about one or two axes of a few degrees, across one boundary while the differences across each of the other two boundaries could be almost the maximum. The conditions of growth and the shape of the specimen caused the boundaries to be perpendicular to the surface of the specimen. The angles between the boundary surfaces were therefore the same as the angles between the traces of the boundaries on the surface. The results of the experiments showed that when the specimens were examined without prior heat treatment the angles between the boundaries were difficult to measure because of their curvature near the common point, but it was estimated that they were always between 100° and 140°. If the specimen was subjected to a temperature a few degrees below its melting point for periods of about a day, it was found that the boundaries had become very nearly straight, and that the angles were always within a degree of 120°.

It is concluded from these observations that the boundaries were able to move at temperatures a little below the melting point, that their movement resulted in a reduction of the total area, and therefore the energy, of the boundaries and that the energies of all the boundaries that were compared were equal although the differences of orientation across the boundaries varied widely from one boundary to another.

These observations, which were made with tin, have not so far been repeated with other metals so it is not possible to generalize the results.

MOTT [47] has pointed out that making certain assumptions based on a transitional lattice theory of the boundary, it can be shown that the energy of the boundary should be independent of the relative orientations of the crystals, unless the difference of the orientations is only a degree or two. Thus observations on the variation of energy with relative orientation do not discriminate between the amorphous cement and the transitional lattice theories, although at first sight it would appear that the experimental results quoted above offer some support to the amorphous cement theory.

The extent of the grain boundary movement occurring during the heat treatment of specimens of the type described above is small. On the other hand, specimens have been produced[1] in which two boundaries between the same two crystals were made to meet at an

acute angle. Some 'rounding off' of the angle occurred while the region was still hot just after solidification, but very considerable shortening of the boundary occurred during subsequent exposure to suitable temperatures for longer times. The orientations of the crystals in specimens of this type were adjusted so that there was no possibility of stresses arising from differential thermal expansions.

It is inferred from these results that crystal boundaries can move as a result of the tendency of the boundary area to become a minimum, presumably as a result of the existence of a finite free energy of the boundary.

The movement of the boundaries in a strain free metal appears to contradict the accepted generalization that grain growth does not occur in a metal 'as cast'. The conditions under which considerable movement took place, however, were artificial, and similar conditions probably occur naturally only in rare instances.

The more general conditions for the movement of crystal boundaries occur when a strained metal is heated. Here the reduction of area of the boundaries is probably not the predominating influence. The reduction in free energy which motivates the occurrence probably arises from the fact that some crystals retain more strain energy than others. A boundary between two crystals with different strain energies will tend to move into the one with the higher strain energy. Thus the less strained crystal 'grows' at the expense of the more strained one. Local fluctuations of the position of the boundary, assumed to be a transition lattice a few atoms thick, would be more probable, and would therefore occur more frequently in the direction in which an atom from the more strained lattice is pulled further from its equilibrium position while an atom from the boundary is allowed to take up its position on the more perfect lattice. The existence of an activation energy for boundary migration supports such a mechanism.

MOTT[38] suggests that the movement of grain boundaries is a result of a process that is essentially the same as that which is responsible for grain boundary shear, namely the local melting of small 'islands' of atoms. This process gives an activation energy of the right order of magnitude. Such an explanation of boundary migration could not be satisfactory in relation to an amorphous boundary layer, but an alternative explanation may be advanced as follows. It is necessary to assume, in order to examine the amorphous layer hypothesis at all,

that the amorphous material is in equilibrium with the crystalline material; otherwise one or the other would disappear, for at any given temperature, if the same amorphous layer were not in equilibrium with both strained and unstrained crystals, it would be more stable than the former and less stable than the latter and consequently would continuously 'dissolve' the more strained crystal while rejecting atoms on to the less strained crystal.

The hypothesis that the amorphous material must be in equilibrium with the crystals at any temperature below the melting point demands that the free energy of the boundary material must vary with temperature according to a different law from that followed by the liquid. Thus some property of the boundary material must change with temperature. This cannot be the structure if this resembles that of the supercooled liquid. It must therefore be some function of the thickness, but since the thickness would have no effect on the free energy of the atoms that are not very close to the limiting surfaces of the boundary, it follows that the boundary must be so thin that the whole of it is affected by the neighbouring crystals. It is difficult to see how a layer of this kind differs from a transitional lattice.

CHEMICAL EFFECTS AT GRAIN BOUNDARIES

Considerable attention has been paid to the chemical effects of grain boundaries. DESCH[48] describes a number of phenomena, and a great deal of information concerning intercrystalline corrosion, both in the absence and in the presence of stress, has been gathered during more recent years.

The literature on this subject is too extensive to be reviewed here but a very brief general picture may be attempted, there being, of course, exceptions to the generalizations made.

Whatever the exact nature of the grain boundary, it is clear that there is a region where the atoms cannot be arranged on a regular crystalline lattice. Consequently it is to be expected that, in general, diffusion will occur at the boundaries more readily than elsewhere. Similarly there will exist, at the boundary, regions where small foreign atoms and others where large foreign atoms will fit more readily than within the crystal lattice of the grains. In the normal process of diffusion, foreign atoms arriving at these positions will tend to remain and the concentration of impurities at the boundary will be higher

than that in the grains themselves. Similarly for precipitation from a supersaturated solid solution it is to be expected that nucleation of precipitates will occur more readily at the boundary than, in general, within the grains. Most of the phenomena observed may be explained qualitatively in terms of these effects. Intercrystalline penetration by other metals,[48] and intercrystalline attack, where the impurities are such as to react naturally with the corroding medium, are readily explained. In the case of intercrystalline attack where precipitation has occurred, and the precipitate is anodic to the grains, electrochemical corrosion explains the behaviour. Where the precipitate is cathodic to the grains, intercrystalline attack is explained by the existence of a region, adjacent to the precipitate, which is depleted of solute and is consequently anodic to both the precipitate and the body of the grain, and which is therefore removed by electrochemical attack.

These phenomena, however, do not throw much light upon the actual arrangement of the atoms at the boundaries, and it is from observations upon pure metals and solid solutions of pure metals that information on this point may be expected. The investigation is complicated by the tendency for insoluble impurities to concentrate at the boundary, as even in metals of high purity the impurity content is sufficient to cause a relatively high concentration in the boundary region if all the impurity is segregated there. In the experimental observations discussed, however, this consideration does not invalidate the conclusions drawn.

Discussing the attack of β brasses by mercury, Desch points out the chemical difference between grains and grain boundaries in apparently homogeneous solid solutions. More lately GRAF[49] suggested that changes in composition near the grain boundaries caused the intercrystalline stress corrosion to be observed in silver-gold solid solutions. In an attempt to investigate whether in fact such a variation in composition does occur in solid solutions, PHILLIPS and BRICK[50] determined by x-ray measurements the lattice parameters of aluminium-copper alloys of various compositions up to a maximum copper content of 5·65 per cent. The copper was retained in solid solution by quenching after homogenizing. Parameters were determined for fine and coarse grained polycrystalline wire specimens and single crystal wire specimens. In addition coarse and fine powder specimens in which the particles were in effect single crystals were examined. It was found that the parameters determined for the single crystal wire and the powder

specimens agreed closely, whereas those for the coarse grained polycrystalline specimens were smaller, and those for the fine grained specimen were smaller still. As the lattice parameter of the aluminium-copper solid solution decreases with increasing copper content, these results could be interpreted as indicating that for any given average composition there was more copper in solution in the body of the grains in the polycrystalline specimens than in the single crystals, and less at the grain boundaries. Investigation of the solid solubility of copper in aluminium supported this observation in that copper appeared to be more soluble in single crystals than in polycrystalline specimens. However, when the actual values are examined such an explanation is impossible, as Phillips and Brick point out. In a polycrystalline alloy of 4 per cent average copper content the measured lattice parameter was the same as that of a single crystal containing 5·5 per cent copper. If this means that the average copper content of the body of the grain in the polycrystalline case was 5·5 per cent while the boundary region was depleted, the boundary region, even if entirely free from copper, would occupy about 27 per cent of the volume of the crystal. It was concluded that the presence of the boundary modified the lattice parameter, but no explanation of this was offered.

DEAN and DAVEY,[61] however, suggested that there could be a marked change in composition near the boundary in solid solutions. They examined the copper content at grain boundaries in zinc-copper alloys of varying composition up to 2·1 per cent copper. The specimens, of fairly large grain size, were coated with a lacquer, and this was cut through at the grain boundaries. Material from the boundaries was removed electrolytically and spectrographically analysed. Material was similarly removed and analysed from different parts of the grains. In all cases the copper content of the material removed from the boundaries was less than that of the material from the body of the grain. The magnitude of the effect observed was surprising. For example, in a specimen of average copper content 1·07 per cent they obtained at the grain boundaries 0·8 per cent copper and in the body of the grain 1·1 per cent. Although the width of the cut in the lacquer is not quoted by Dean and Davey it must clearly have been many atomic diameters, yet the results indicate that the composition of the grain was affected by the boundary for a distance at least comparable to this. It does not appear likely that such changes could be produced by a simple transition at the boundary from one lattice orientation to another.

Observations on precipitation, however,[52] suggest that the boundary structure is dependent on the relative orientations of the grains and the boundary direction, as would be expected in accordance with a transition picture. It was observed in copper-beryllium alloys that there were pronounced and abrupt changes in the precipitation at a grain boundary when there was a change in relative orientation where twinning occurred in one of the grains, and in an aluminium-magnesium alloy the degree of precipitation between two grains depended upon the orientation of the boundary itself. These observations were explained by suggesting that nucleation for precipitation should be easier in those cases where there was more irregularity at the boundary due to difficulty of transition from one crystal lattice to the other.

Similar support is given to the transition from one orientation to another by LACOMBE and YANNAQUIS,[53] who observed that when super-purity aluminium was attacked by a 10 per cent solution of hydrochloric acid there was generally a rapid attack of the boundaries, but the attack depended upon the relative orientation of the grains. Boundaries where the two grains were of substantially the same orientation and those where the grains were in twin relationship were particularly resistant to attack. From this observation Lacombe and Yannaquis suggested that the intercrystalline attack must be attributed to the discontinuities at the junction of adjoining lattices, the discontinuities being greater the greater the difference in orientation of the grains. It is suggested that even with aluminium of such purity (99·9986 per cent) there might have been sufficient impurity present at the boundaries to effect the observed result. The variation with the nature of the boundary, however, would still be significant, since it would indicate that the concentration of impurity was greatest where the transition from one lattice to the other was most difficult.

In the light of the evidence, both theoretical and experimental, it may be concluded that the boundary between two crystal grains represents a transition from one crystal lattice to another. There is no reason to believe that there exists at the boundary a distinct amorphous phase. The indications which appear to point to the existence of such a structure may all be accounted for in terms of a transition region in which the atoms are situated at defined mean points dictated by the forces exerted by the crystals, although small regions may become temporarily disordered under the action of thermal fluctuations.

References

[1] Chalmers, B. *Proc. roy. Soc. A* 196 (1949) 64
[2] Quincke, G. *Berliner Monats.* (1868) 139; *Proc. roy. Soc. A* 76 (1905) 431
[3] Brillouin, M. *Ann. Chim. Phys.* 13 (1898) 377
[4] Ewing, J. A. and Rosenhain, W. *Phil. Trans. roy. Soc. A* 195 (1901) 279
[5] Beilby, G. T. *Proc. roy. Soc.* 72 (1903) 218; *J. Inst. Met.* 6 (1911) 5
[6] Bengough, G. D. *J. Inst. Met.* 7 (1912) 123
[7] Rosenhain, W. Discussion of Bengough's paper *Ibid* 7 (1912) 178
[8] —— and Ewen D. *J. Inst. Met.* 8 (1912) 149
[9] —— and Humfrey, J. C. W. *J. Iron Steel Inst.* 87 (1913) 219
[10] —— and Ewen, D. *J. Inst. Met.* 10 (1913) 119
[11] —— and Archbutt, L. *Proc. roy. Soc. A* 96 (1919) 55
[12] Jeffries, Z. *J. Amer. Inst. Metals* 11 (1918) 300; *Trans. Amer. Inst. min. met. Engrs* 60 (1919) 474
[13] Rosenhain, W. Metallurgist, suplt. *Engineer* 1 (1925) 2; *ibid* 144 (1927) 422
[14] Gough, H. J. *Trans. Faraday Soc.* 24 (1928) 137
[15] Hargreaves, F. and Hills, R. J. *J. Inst. Met.* 41 (1929) 257
[16] Townsend, J. R. *Proc. Amer. Soc. Test Mat.* 27 (ii) (1927) 153
[17] Bragg, W. L. *Proc. phys. Soc.* 52 (1940) 105
[18] Cottrell. This volume p 77
[19] Burgers, J. M. *Proc. phys. Soc.* 52 (1940) 23
[20] Carpenter, H. C. H. and Elam, C. F. *Proc. roy. Soc. A* 100 (1921) 329
[21] Aston, R. L. *Proc. Camb. Phil. Soc.* 23 (1926) 549
[22] Barrett, C. S. and Levenson, L. H. *Trans. Amer. Inst. min. met. Engrs* 137 (1940) 112
[23] Chalmers, B. *Proc. roy. Soc. A* 162 (1937) 130
[24] Boas, W. and Hargreaves, M. E. *Ibid A* 193 (1948) 89
[25] Chalmers, B. *Proc. phys. Soc.* 52 (1940) 127
[26] Taylor, G. I. *J. Inst. Met.* 62 (1938) 307
[27] Hanson, D. and Wheeler, M. A. *Ibid* 45 (1931) 229
[28] McKeown, J. *Ibid* 60 (1937) 201
[29] Greenwood, J. N. and Worner, H. K. *Ibid* 64 (1939) 135
[30] Moore, H. F., Betty, B. B. and Dollins, C. W. *Univ. Illinois Bull.* (1935) 272
[31] King, R., Cahn, R. W. and Chalmers, B. *Nature, Lond.* 161 (1948) 682
[32] Zener, C. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 155
[33] Barnes, A. and Zener, C. *Phys. Rev.* 58 (1940) 87L
[34] Zener, C., Van Winkle, D. and Nielsen, H. *Trans. Amer. Inst. min. met. Engrs* 147 (1942) 98
[35] Kê, T. S. *Phys. Rev.* 71 (1947) 533; 72 (1947) 41
[36] Zener, C. *Ibid* 60 (1941) 906
[37] Polyak, E. V. and Sergeev, S. V. *C. R. Acad. Sci. U.R.S.S.* 30 (1941) 137
[38] Mott, N. F. *Proc. phys. Soc.* 60 (1948) 391
[39] Chalmers, B. *Proc. roy. Soc. A* 175 (1940) 100
[40] Chaudron, G., Lacombe, P. and Yannaquis, N. *C. R. Acad. Sci., Paris* 226 (1948) 1372
[41] Rosenhain, W. and Humfrey, J. C. W. *Proc. roy. Soc. A* 83 (1910) 200
[42] Humfrey, J. C. W. *Carnegie Mem. Iron and Steel Institute* 4 (1912) 8
Rawdon, H. S. and Scott, H. *Trans. Amer. Inst. min. met. Engrs* 67 (1922) 414
Hemingway, E. H. and Ensminger, G. R. *Ibid* 67 (1922) 392
[43] Fonda, G. R. *Phys. Rev.* 21 (1923) 343
[44] Carpenter, H. C. H. and Elam, C. F. *J. Inst. Met.* 24 (1920) 83
[45] Boas, W. and Honeycombe. *Proc. roy. Soc. A* 186 (1946) 57
[46] Chalmers, B., King, R. and Shuttleworth, R. *Ibid* 193 (1948) 465

[47] MOTT, N. F. Private communication, 1948
[48] DESCH, C. H. *The Chemistry of Solids* London, 1934 p 65
[49] GRAF, L. *Z.W.B. Forschungsbericht (FB)* No. 1954 (1944)
[50] PHILLIPS, A. and BRICK, R. M. *Journal Franklin Inst.* 215 (1933) 557
[51] DEAN, G. R. and DAVEY, W. P. *Trans. Amer. Soc. Met.* 26 (1938) 267
[52] FORSYTH, P. J. E., METCALFE, G. J., KING, R. and CHALMERS, B. *Nature, Lond.* 158 (1946) 875
[53] LACOMBE, P. and YANNAQUIS, N. *Métaux et Corrosion* 22 (1947) 35

4

AGE HARDENING OF METALS

G. C. Smith

DURING THE years 1903–11, the German scientist WILM discovered the alloy now known as duralumin. The unusual properties of this alloy promoted subsequent work upon the ageing of metals. The first detailed work was that of MERICA, WALTENBERG, and SCOTT in 1919, and since then much literature has appeared on the subject.

The knowledge which has been accumulated deals not only with duralumin but also with many other alloys, which like duralumin can be strengthened by an ageing treatment. Theory and practice alike have received detailed study, but the following article deals only with the theory of the ageing process, tracing its development since the earliest work.

WILM[1,2] in his investigations developed a duralumin with the approximate composition: copper 3 to 4·5 per cent, magnesium 0·4 to 1·0 per cent, manganese 0 to 0·7 per cent, silicon 0·3 to 0·6 per cent, iron 0·4 to 1·0 per cent, and the balance aluminium. He found that if a sample of such an alloy was quenched from high temperatures below its melting point and then allowed to stand at room temperature the hardness increased with time. It was found later that this increase in hardness could occur on standing at other temperatures, but, as will be seen later, the course of the hardening differed at different temperatures. This process has come to be known as ageing, and the change in hardness as age hardening. Though other properties of the alloys are changed by ageing, Wilm was concerned mainly with age hardening. Reference to other changes will be made later.

Merica, Waltenberg, and Scott[3] investigated in more detail the effect of varying the heat treatment of duralumin on the properties produced by ageing. Such variables as quenching temperature and ageing temperature were studied. It had been noted previously by BLOUGH that the quenching temperature exerted a considerable effect upon the hardening. Wilm used only one quenching temperature, about 450°C, whereas Merica, Waltenberg, and Scott found that the best mechanical properties were developed in duralumin by quenching

from the highest possible solution temperature that could be used without melting the eutectic. This temperature was taken to be around 520°C. The rate of hardening was found to increase as the ageing temperature increased, and the maximum hardness obtained over a period of ageing of fifteen days was found to occur at temperatures above

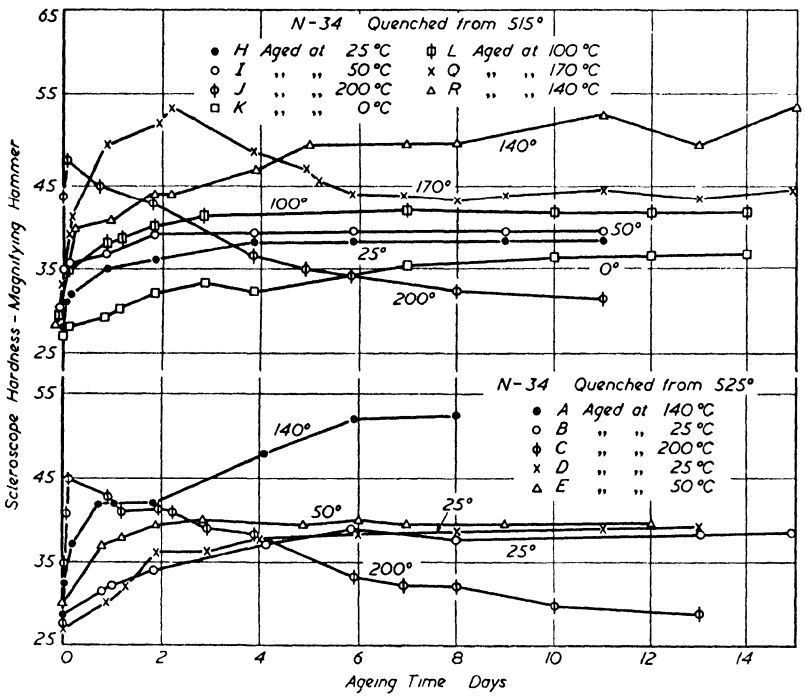


Figure 1. Effect of ageing at different temperatures on the scleroscope hardness of samples quenched from 515°C and 525°C (Merica, Waltenberg, and Scott)

100°C but below about 200°C. At 200°C and above the hardness/time curve passed through a maximum, so that the alloy softened at the longer times of ageing (*Figure 1*). At the lower temperatures it appeared that the hardness became more or less constant after the initial increase.

The mechanical properties investigated in this early work were tensile strength, elongation, and hardness. It was found that the extent to which good ductility could be combined with good strength was dependent to some extent on the temperature of ageing.

EARLY THEORIES

No theory had then been put forward to explain the changes occurring during ageing. Merica, Waltenberg, and Scott considered that all the available evidence seemed to indicate that the hardening was accompanied by a phase change in the alloy, or at least by profound alterations in the atomic or molecular structure. The solid solubility of copper

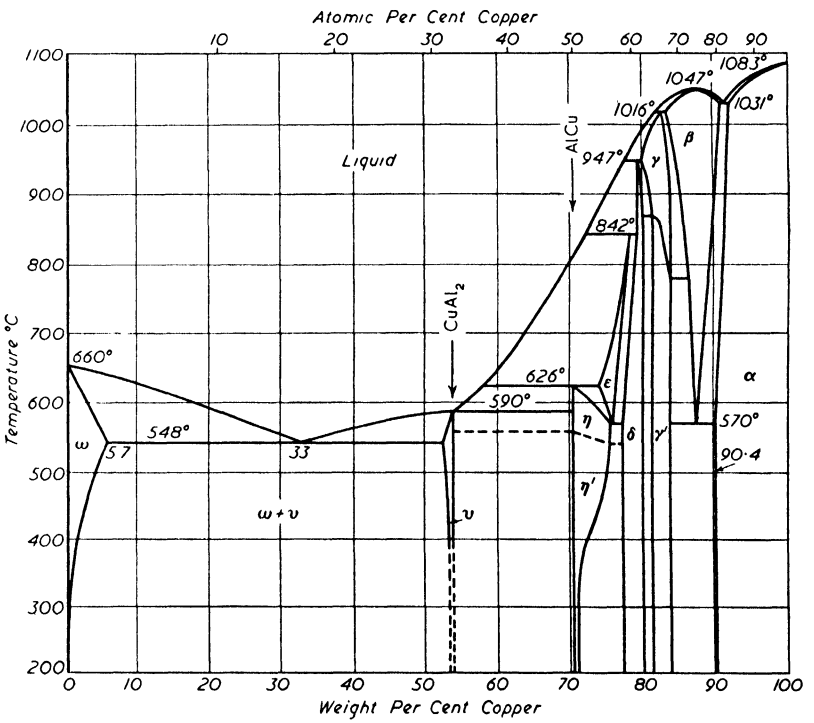


Figure 2. The aluminium-copper equilibrium diagram (After Hansen)

in aluminium was thought to be a factor of great importance. It had been shown previously that the solid solubility of copper in aluminium decreased rapidly with decreasing temperature[4] (*Figure 2*). On slow cooling of an aluminium-copper alloy containing four per cent copper, precipitation of the intermediate compound $CuAl_2$ should therefore take place from the uniform solid solution which existed at the solution treating temperature. This would mean that the structure obtained in the alloy at room temperature would consist of $CuAl_2$ particles embedded in a matrix of the solid solution of copper in aluminium, of concentration

predicted from the equilibrium diagram at room temperature. Should the alloy be cooled rapidly from the solution temperature, however, as in quenching, the precipitation of the $CuAl_2$ would be suppressed, and there would be retained at room temperature a supersaturated solid solution of copper in aluminium. It was believed that the subsequent decomposition of such a solid solution led to the changes observed on ageing.

The authors pointed out that if the solid solution was aged at the temperature of liquid air (— 180°C) no hardening took place. They attributed this to the fact that although the alloy was not in equilibrium the rate of formation of $CuAl_2$ was then so slow that no precipitation could take place. At room temperature or above the mobility of the atoms would become great enough to enable precipitation of $CuAl_2$ to take place and result in the formation of a very fine dispersion of precipitate which was assumed to be colloidal. It was this fine precipitate which was proposed as the important factor causing age hardening. In support of this, experiments were carried out to determine whether a phase change occurred during ageing.

It was shown by means of thermal analysis that an evolution of heat took place in the range 250 to 275°C on heating a specimen of duralumin that had been quenched but not aged. The heat evolution was taken to indicate the formation of a stable phase from an unstable phase. No such evolution was found in a specimen which had previously been slowly cooled from 520°C. Furthermore a specimen that had been quenched and then aged at 100 to 150°C to obtain maximum hardness showed little or no evolution of heat upon subsequent heating. Thus, whatever change took place in a quenched specimen causing heat evolution on subsequent heating, it must have taken place during the time required for ageing in a specimen that had been quenched and aged, *i.e.* the time during which the hardening of the alloy had taken place. These results seemed therefore to indicate that a phase change did in fact take place during ageing.

In an endeavour to check this more definitely, the microstructural changes which took place on ageing were also investigated, since if ageing was due to the decomposition of a supersaturated solid solution, with resultant precipitation of $CuAl_2$, some change in the microstructure could be expected. This might take the form of a visible precipitate of $CuAl_2$, but there was a possibility that the precipitate

might be submicroscopic. Should this be so, however, the presence of a very fine precipitate in the alloy might reasonably be expected to exert some effect upon its etching characteristics. It was pointed out that the etching properties of tempered martensite were known to be different from those of martensite itself, a fact explained by the formation of a very fine dispersion of Fe_3C from the martensite on tempering. Specimens of duralumin which had been quenched and immediately etched were compared with specimens that had been quenched and then aged at 130°C in order to develop maximum hardness, but no difference in structure was observed between the specimens, either in the general appearance or the amount of precipitate present. In the alloys used it was found that on solution treatment all the $CuAl_2$ did not dissolve, and so some was always present in the as quenched structure. Particles of $FeAl_3$ were also present and a constituent which was thought to contain iron, silicon, and aluminium. The failure to observe a change in microstructure during ageing meant, therefore, that Merica, Waltenberg, and Scott had no direct evidence of precipitation. They pointed out, however, that a difference in the general rate of etching between the quenched and quenched and aged states might be obscured by the presence in the specimens of some other fine dispersion of material, which might perhaps be the constituent which was thought to contain iron, silicon, and aluminium, called the X constituent. Thus a microscopical examination of duralumin made from pure aluminium, free from silicon and iron, might have yielded better results. It was also found that even after annealing an alloy containing 3 per cent copper for 20 hours at 300°C no precipitate was visible, after a treatment at 500°C which had resulted in the solution of all the $CuAl_2$. From this it was concluded that the velocity of coalescence of the $CuAl_2$ was low.

The conclusion that the hardening constituent was $CuAl_2$ alone of all the constituents present in the duralumin was reached as follows. It was known that no phase change took place in pure aluminium on heating from room temperature up to the melting point. The heat evolution which had been observed in duralumin around 250°C to 300°C was taken, therefore, to be associated with some precipitation process. The possible constituents which could precipitate were $FeAl_3$, the X constituent, $CuAl_2$,Mg_4Al_3, or Mg_2Si. Aluminium containing the same amounts of $FeAl_3$ and the X constituent as duralumin

were known not to give any hardening effects or to show thermal effects upon heating, thus excluding them as the cause of hardening. Furthermore, alloys which contained magnesium in amounts up to three per cent did not harden on ageing. Only $CuAl_2$ remained as the possible cause, although the consideration given to Mg_2Si seems to have been rather brief.

By attributing the ageing effects to the precipitation of $CuAl_2$ in a fine form, many of the known but puzzling facts about ageing could be explained. The dependence of the extent of hardening upon the solution temperature was explained by the greater solubility of copper in aluminium at higher temperatures. More copper was thus taken into solid solution in the aluminium, and on subsequent ageing a greater amount of $CuAl_2$ would be precipitated, with the development of a greater hardness. The shape of the time/hardness ageing curves which at the higher ageing temperatures showed an initial increase in hardness, followed by a decrease, was attributed to a variation in size of the $CuAl_2$ particles. Initially the precipitate particles would be of an extremely small size, but at fairly high ageing temperatures growth and coalescence would take place. It was assumed that there was a critical size of particle which would produce the maximum hardening effect and that this size was larger than the size of particles formed on first precipitation. Consequently, as growth to the critical size took place, the hardness increased, reaching a maximum when the critical size was attained and thereafter decreasing as the particles grew still larger. The critical size was assumed to be submicroscopic, which would then explain the absence of any microstructural evidence of precipitation during hardening.

The mechanism whereby the fine precipitate caused an increase in strength was attributed by Merica, Waltenberg, and Scott to surface tension forces which would extend the influence of a precipitate particle beyond its actual size, and thus exert a strengthening effect upon the lattice in those regions where it was present. The smaller the precipitate particles the greater would be the volume of the matrix lattice affected by the surface tension forces, so that a greater strengthening of the alloy would result.

Discussing Merica, Waltenberg, and Scott's paper, however, JEFFRIES advanced another theory to explain the strengthening effect, a theory later amplified by JEFFRIES and ARCHER.[5] They pointed out

that by means of the fine dispersion of $CuAl_2$ in duralumin it was possible to increase the hardness of pure aluminium by approximately ten times. They reasoned that such an increase in strength must arise from an increase in the inherent cohesion of the aluminium lattice since it was inconceivable that the presence of small, hard, disconnected particles of $CuAl_2$ could exert any specific effect by themselves. It was assumed that the increase was accomplished by strengthening the weakest regions of the lattice. These were taken to be the slip planes, which were known to be the regions where deformation of material took place. Jeffries and Archer postulated a 'keying action' upon the slip planes as the mechanism by which the small particles were supposed to act. The presence of the particles was assumed to make the slipping of the planes over one another more difficult, thus increasing the mechanical strength of the alloy. The larger the number of particles present, then the greater the hardness and strength developed owing to increased keying action. It was considered that the key formed by the little group of aluminium atoms which would surround a copper atom when the copper was in solid solution would be more easily sheared through than a discrete particle of crystalline $CuAl_2$. The conclusion was therefore drawn that the maximum hardness would be obtained when there was present in the alloy a dispersion in which each particle was of the smallest size possible, consistent with the fact that it retained the crystallinity characteristic of the $CuAl_2$ compound. With a given amount of precipitate, larger particles would have less strengthening effect because there would be fewer keys. This would explain why at some ageing temperatures the hardness increased initially and then decreased as the particles coalesced beyond the critical dispersion size.

Merica[6] criticized this theory on the grounds that if a given amount of hard particles dispersed in an alloy was considered, then the same shearing area would be exposed by the particles on the slip planes of the matrix, independent of their actual size, and the keying action should therefore be independent of the size of the dispersed particles.

Jeffries and Archer disagreed with this. They thought that the keying action should not be taken as proportional to the area of the slip planes which the particles intersected since the hard particles were supposed never to be sheared through themselves. Hence the presence on a slip plane of a hard particle of any size would prevent slip upon that particular plane. The finer the precipitate became,

therefore, the greater would be the number of planes of the matrix that were keyed, so that the strength would increase, despite the fact that precipitation of the solute element from solid solution would cause the inherent resistance to slip of the matrix to be diminished. The keying action set up by precipitate particles would more than compensate for the decrease in inherent resistance involved in the precipitation.

ROSENHAIN introduced a different view. He interpreted hardness and strength, *i.e.* resistance to deformation, as a function of the amount of distortion existing in the lattice of a metal or alloy and considered that the hardening of duralumin during ageing was due to an increase in disorder and distortion in the lattice of the matrix brought about by precipitation of the $CuAl_2$. This idea was criticized by Jeffries and Archer. According to Rosenhain, the hardness of a solid solution was attributable to lattice distortion. It was therefore to be expected that as the quantity of solute present in a solid solution increased, so would the hardness, since the distortion present in the lattice would be increasing. Consequently any process which resulted in a diminution of the concentration of solute in the solid solution would result in a decrease in hardness by reducing the lattice distortion, and precipitation of $CuAl_2$ should therefore cause softening, not hardening. In support of this it was pointed out that when duralumin was over aged it was markedly softened due to the extensive rejection of atoms from the solid solution. This being so, there was no reason why the softening should not take place from the moment that precipitation began, since the reduction of lattice distortion would be taking place, and it was therefore necessary to assume that some other factor such as the formation of particles of a critical size, caused the initial increase in hardness observed on ageing.

Except for one or two isolated experiments there was little early work carried out upon the changes of physical properties other than those of strength, brought about by ageing. In 1922, however, KONNO[7] in Japan and FRAENKEL and SCHEUER[8] in Germany followed the variation of some of the electrical properties of duralumin, and found them to be dependent on the temperature at which the ageing was carried out. At room temperature the resistance increased as the strength increased, but at higher temperatures the resistance sometimes decreased while the strength was increasing. These facts were very difficult to explain upon the critical dispersion theory, since it was

generally considered that the precipitation of a substance from solid solution would result in a decrease in the electrical resistance. The discovery that the resistance could increase at the same time as the hardness seemed to mean, therefore, that the hardening effect could not be caused by a process that involved simple precipitation.

In consequence the theory of ageing of duralumin was modified by certain workers[9] who postulated the presence in the alloy of two solid solutions of different resistivity and hardness, the slow transformation of one into the other leading to the characteristic changes during ageing. It was assumed by Honda that Mg_2Si was essential to these changes, the varying solubility of Mg_2Si in what was termed $CuAl_2$ solid solution being the governing factor. The Mg_2Si was thought to be more soluble in the $CuAl_2$ solid solution at higher temperatures and to separate on cooling, *via* an intermediate solid solution. Thus, if the solid solution at high temperatures was known as *A*, and the equilibrium state at room temperature, *i.e.* when Mg_2Si particles were mixed with the $CuAl_2$ solid solution, known as *C*, then it was proposed that the transformation of *A* to *C* normally took place through an intermediate solid solution *B*, of Mg_2Si in the $CuAl_2$ solid solution. *B* was supposed to be very hard compared with the *A* or *C* states, and it was assumed that *A*, *B*, and *C* had different resistivities. Normally on slow cooling the changes took place in the order *A* to *B* to *C*, so that the final product was *C*, and the alloy was fairly soft. On quenching, however, the transformation of *A* was suppressed and this was therefore retained at room temperature. It would, however, be unstable and tend to break down to the final stage *C*, but in so doing was assumed to pass through the hard intermediate stage *B*. This would therefore lead to the increase in hardness and resistance observed on ageing at room temperature. The changes were assumed to take place only very slowly at room temperature, so that the hard *B* state would remain for a considerable length of time. At higher temperatures, however, the final transformation into state *C* would take place so that after the initial increase in hardness the alloy would eventually soften and also change in resistance. The actual way the resistance changed at any temperature would be governed by the amounts of *A*, *B*, and *C* present at various times during the ageing.

That the above theory involved Mg_2Si as an important factor was of some interest, for in their original theory Merica, Waltenberg, and

Scott had attributed the hardening during ageing entirely to the precipitation of $CuAl_2$ and had assumed that any Mg_2Si present in the alloy had no influence.

In connection with the extent to which Mg_2Si and $CuAl_2$ individually influenced the ageing of duralumin some early work of GAYLER[10] should be mentioned. She concluded that the ageing could not be said definitely to be due to one or the other of the two constituents. The evidence available indicated however that the predominant constituent with regard to room temperature ageing was Mg_2Si.

THE PRECIPITATION THEORY

Gayler also studied the changes which took place when an alloy that had been aged to maximum hardness at room temperature was subsequently held at a more elevated temperature. From this an important point arose. It was found that the hardness could in some cases suffer an initial decrease at the higher temperature, and then subsequently increase again. This was thought to be associated with the manner in which $CuAl_2$ and Mg_2Si tended to come out of solid solution. It was suggested that the changes which took place during ageing at room temperature were due to the precipitation of Mg_2Si, while the changes on further heat treatment at a higher temperature were due to the precipitation of $CuAl_2$. The initial softening produced on heating at the higher temperature after the preliminary ageing at room temperature would be due to the coalescence of precipitated Mg_2Si particles, but subsequently the precipitation of $CuAl_2$ would be effective in causing another increase in hardness. The high temperature necessary for the precipitation of the $CuAl_2$ was thought to be due to the small difference in solubility of copper at high and low temperatures in the presence of magnesium and Mg_2Si.

In the ageing of duralumin the ageing effects were identified with a change in solid solubility of a constituent with temperature. Aluminium-zinc alloys represented another type of system which, it was known, could exhibit ageing effects, and a certain amount of early work[11,12] was done on such alloys (*Figure 3*). The essential feature here is the suppression of a eutectoid change by quenching, and hence the retention at room temperature of the solid solution which is normally stable at elevated temperatures. This would be the β phase in the case of the aluminium-zinc alloys. Subsequent ageing of this phase, resulting in breakdown to $\alpha + \gamma$, was shown to cause changes in

properties resembling those taking place during the ageing of aluminium-copper alloys. It was in connection with the hardening shown by these alloys that ROSENHAIN[13] extended his previous idea of the strain in the lattice as the important factor by proposing that the formation of precipitate particles left the matrix lattice in a practically amorphous condition, which gave the lattice an increased resistance to deformation.

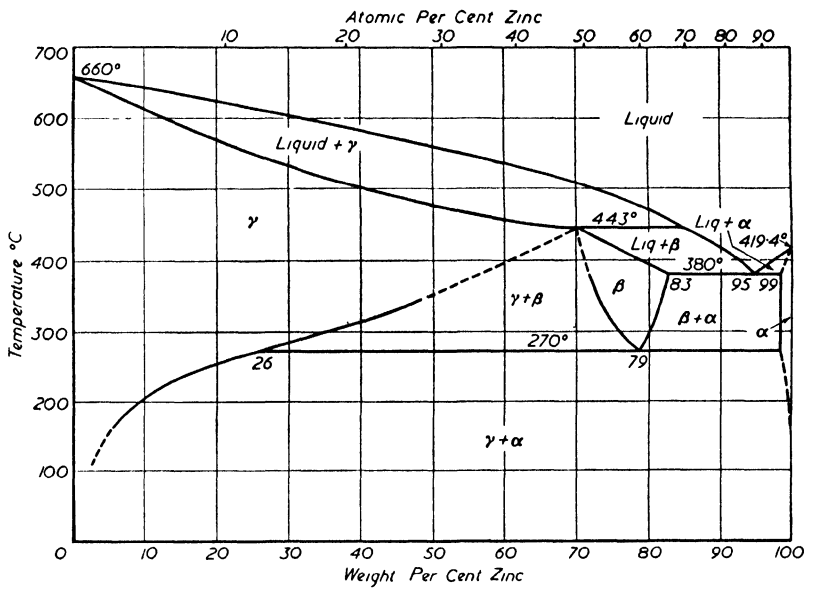


Figure 3. The aluminium-zinc equilibrium diagram (After Hansen)

The suggested mechanism of hardening by a fine dispersion of precipitate attracted a considerable amount of interest, and other theories to explain hardening were put forward. Colloidal behaviour was regarded by some as being important, and according to its supporters it could explain all the known facts about ageing. This theory was based on the fact that the properties of a dispersed system change periodically with the degree of dispersion, from which it was concluded that there was a 'maximum degree of colloidality' at which a system exhibited colloidal phenomena to the greatest extent, a certain degree of dispersion being considered critical. Size of particles alone was the important factor, and shape, internal structure, or orientation were not considered to have any effect. This was considered to be a strong point in favour of the colloidal theory, as the Jeffries and Archer theory of slip interference had been criticized on the grounds that the

keying particles would be expected to be inherently hard. It was pointed out, however, that a finely divided dispersion of a weak constituent was in certain cases capable of producing a considerable degree of hardening.

There was as yet no direct evidence that precipitate particles did in fact form during the early stages of ageing, though it was known that they appeared after long ageing times when softening had set in. The techniques employed in following ageing may perhaps have been responsible for this, as little early work seems to have been carried out in investigating ageing changes microscopically or by x-rays. The majority of the early workers appear to have been concerned mainly with the changes in mechanical properties and electrical resistance.

Sachs,[14] however, found for duralumin that aged crystals were etched more quickly than annealed crystals. He also took x-ray photographs of aged duralumin but found no evidence in the photographs of any internal stresses. He pointed out that such internal stresses were known to alter the Laue photographs, and deduced that the absence of any alteration meant that there were no internal stresses.

Further x-ray examination in 1929 by GÖLER and SACHS[15] showed that in single crystals of a 5 per cent copper-aluminium alloy ageing could be detected by means of tensile tests before any disturbance of the quenched solid solution could be found by x-rays. It was therefore concluded that there was a type of ageing that could proceed without any detectable change in the solid solution lattice, and more particularly without any detectable precipitation of a constituent from solid solution. This meant that if precipitation was taking place it was very limited in amount and finely dispersed.

An extensive investigation into ageing, which revealed much new information, was carried out in 1929 by GAYLER and PRESTON,[16] who followed the changes in hardness, density, electrical resistance, and made x-ray measurements. They concluded that the changes in density and lattice parameter which took place during ageing at room and higher temperatures indicated that precipitation from solid solution was taking place. It was shown that at room temperatures hardness and electrical resistance both passed through a maximum, although at different times. With resistance the maximum was reached in from 1 to 3 hours but with hardness in from 2 to 3 days. At higher temperatures of ageing, such as 200°C, the maxima coincided. The x-ray measurements showed that besides a change in lattice parameter the crystals

appeared to be in a disturbed state, which was gradually relieved by further ageing treatment at higher temperatures. This was in contradiction to Sachs' earlier experiments. Gayler and Preston concluded that the precipitation of solute from solid solution during ageing took place in two stages, first the rejection of atoms of the dissolved metal from the lattice, a process which was thought probably to be accompanied by the formation of molecules, and secondly the coagulation of the rejected atoms and molecules to form minute crystallites. The two stages were considered to overlap somewhat and to have different effects upon the properties of the material. The first stage was thought to entail a considerable disturbance of the lattice, producing the increase of electrical resistance observed during the initial stages, and causing the broadening of the x-ray lines which occurred at the same time. The second stage, coagulation, was thought to be a softening process whereby the lattice strain was gradually relieved so that the resistance, and eventually the hardness, began to decrease and at the same time the lines on the x-ray spectrum became less diffuse.

Gayler and Preston also commented upon a problem of ageing which has still not been fully investigated, namely the intensity of the hardening effect. It was known that the extent of the hardening varied considerably from system to system and appeared to depend to a certain extent upon the nature of the precipitating phase. Thus, if the hardening was due to the precipitation of a pure metal, it did not appear to be as intense as that which occurred when a compound was precipitated. For example, the copper-iron alloys did not harden as intensely as the aluminium-copper alloys. This was explained by assuming that the separation of a pure metal from solid solution did not entail as much distortion of the lattice as the separation of a compound of the metal, since in the former case no molecule formation took place. A more intense hardening would occur when the compound was formed by the combination of two or more solute elements with the solvent than when only a single solute element combined with the solvent. Thus a greater percentage increase in hardness was in fact found when an alloy which precipitated Mg_2Si was aged, compared with an alloy which precipitated $CuAl_2$.

GWYER[17] in 1929 and GAYLER[18] in 1930 reviewed the state of the theory of ageing. Both pointed out that the precipitation theory had not met with universal acceptance, as it seemed unable in its simplest form to explain some of the changes known to occur. Gayler

mentioned that the two main objections to the theory at the time of writing were the lack of any microscopical evidence of precipitate particles during ageing and the anomalous changes in the electrical resistance, *i.e.* the initial increase which took place under certain conditions. Gayler mentioned that recent work which gave definite microscopical evidence of the presence of $CuAl_2$ during the ageing of a copper-aluminium alloy had been obtained, but made no reference to the paper in question. In connection with the resistance changes, she pointed out that the x-ray spectra became diffuse during ageing and that if this indicated the strain existing in the lattice the initial increase in resistance would be explained, assuming that the resistance was dependent upon the lattice strain. Little differentiation was made, however, between the strain causing electrical resistance changes and that causing hardness changes, and it was apparent that since the hardness and resistance did not always change together there must be some difference between the two. Gayler mentioned that there was no direct x-ray evidence of the precipitation of compounds from solid solution during ageing. This was assumed to be due to the difficulties of detecting very small percentages of separating phases by such methods. For copper-beryllium alloys, however, it had been found possible to detect the lines of the separating phase. These lines were diffuse, which indicated that the initial size of the particles was very small. Despite the above anomalies, however, Gayler concluded that the facts then established gave strong confirmation of the idea that hardening during ageing was due to the precipitation of highly dispersed particles.

LATTICE DISTORTION THEORIES

It would appear that the effect of these particles was attributed largely to the strain which their formation generated in the lattice. The increase in hardness due to ageing was thus linked with the increase in hardness due to cold working, which was also thought due to lattice distortion. Certain workers, however, still favoured the keying mechanism proposed by Jeffries and Archer.

A prominent advocate of the lattice strain idea was ROSENHAIN[19] who explained by this theory the phenomena observed when material was aged first at low temperatures and then at higher temperatures. (These have since been termed retrogressive phenomena.) It was known that under such circumstances an initial hardening took place at

room temperature and then, on raising the ageing temperature, a softening occurred which was later followed by a fresh hardening. In Rosenhain's opinion this could be explained by referring to the two stages which Gayler and Preston postulated, namely initial precipitation of the solute in atomic or molecular form, followed by coagulation. If it is assumed that the first stage brought about the extra distortion of the lattice which caused hardening, while the second stage eventually caused a softening by relieving the distortion, then a net hardening of the alloy would necessarily be due to a predominance of the first stage, and a net softening to a predominance of the second. Rosenhain also assumed that the strain which was set up in the lattice by the first stage inhibited the process of diffusion, which had to take place in order that either of the above stages could occur. Thus the initial strain set up by precipitation would inhibit diffusion, making precipitation more and more difficult until it stopped altogether. The second coagulation stage would also cease, since this also would depend on diffusion. Thus a stable state could be reached while the matrix was still supersaturated with respect to the solute atoms. If, however, the ageing temperature was raised, further diffusion would become possible. The initial precipitation would be resumed until a new stable state was reached at the new ageing temperature, due to diffusion being once more inhibited. This would thus lead to a new hardening. The fact that the new hardening was in general preceded by an initial decrease was explained by assuming that the first effect upon raising the temperature was to accelerate the coagulation process so that a net softening would occur, and hardening would not begin until the first initial precipitation process became appreciable. This explanation of the retrogression effects did not, therefore, make the assumption that two different ageing mechanisms were in operation at the two temperatures involved. Advocates of this latter viewpoint thought that in the case of aluminium-copper alloys the precipitation of $CuAl_2$ was responsible for ageing effects at one temperature and the precipitation of Mg_2Si at other temperatures. Rosenhain did not reject this idea altogether and thought that it might well be the true mechanism of retrogression effects in the case of duralumin, but it could not explain the effects observed in the case of aluminium containing only the one hardening component, copper.

The importance of lattice distortion in hardening was also stressed by KOKUBA and HONDA,[20] who discarded the idea of precipitation altogether,

since it could not explain all the known facts, and advanced a theory which attributed the changes on ageing to lattice distortion in the supersaturated solid solution. TAMMANN[21] also thought that ageing was due to internal changes distinct from precipitation and that a satisfactory explanation could be based upon the idea that the surplus dissolved atoms in the solid solution accumulated upon distinct planes and lines of the lattice, thus increasing the material's resistance to glide. Other properties, it was reasoned, would also alter, owing to the change in the distribution of the atoms in the lattice. Thus the process of precipitation and formation of crystallites would be associated with a decrease in the resistance of the material to gliding.

The number of papers dealing with ageing phenomena increased rapidly and dealt with a great variety of topics. New systems which exhibited ageing effects were discovered, and the effects of added elements, mechanical deformation before and after ageing, and different heat treatments, upon the properties of the already known age hardening alloys received a considerable amount of attention.

THE KNOT THEORY

The subject was thus already becoming rather complex, and none of the theories put forward to explain the various changes were really satisfactory. However, MERICA[22] in 1932 made an extensive review of the whole field and considered the theoretical aspects of the subject in the light of more recent information, which necessitated modification of the original precipitation theory. The chief anomaly which he cited as being inexplicable in the idea of simple precipitation was that ageing could take place in copper-aluminium alloys at temperatures below 150°C without the appearance of a visible precipitate of $CuAl_2$, or even a submicroscopic precipitate that could be identified by some other experimental method. It was known that ageing could be complete before visible precipitation occurred, and Merica therefore postulated that ageing could occur either by the precipitation of very fine particles or alternatively by structural alterations in the solid solution other than those of actual precipitation, the former taking place at elevated temperatures and the latter at lower temperatures. This seems strange, as it would appear that there was no reason why the same process should not have been considered operative at the higher temperatures as at the lower temperatures. The fact that at the high temperatures particles of precipitate were visible would be merely

a secondary factor, and the actual changes responsible for the ageing phenomena would be the same as at low temperatures, *i.e.* the changes taking place in the lattice prior to actual precipitation.

In the idea which Merica developed to explain low temperature ageing, however, he assumed from the work of HENGSTENBERG and WASSERMANN[23] that diffusion took place during this process. The basis for this assumption was that the above workers had found during ageing a change in the intensity of the x-ray line spectra which resulted in an increased line intensity and a decrease in the intensity of diffuse radiation. On the assumption that diffusion did occur, Merica postulated for aluminium-copper alloys that copper atoms concentrated at those positions in the solid solution lattice which they would occupy just prior to atomic rearrangement and precipitation as crystalline particles of $CuAl_2$. Other positions might have been possible such as the formation of an interpenetrating lattice,[21] or diffusion of the atoms towards positions which they would occupy on the lattice of some unstable molecule or compound, but Merica favoured the simplest hypothesis. The localized concentrations of copper atoms which would arise in the lattice at certain points he called knots, and thus the so-called knot theory of ageing arose. The knot theory was used to explain ageing changes which took place when no experimental evidence of actual precipitation was available, such as with duralumin aged at around room temperatures. The knots were not regarded as separate particles but were assumed to be continuous with the solid solution, so that there was no interface between the knot and the solid solution. The size and shape of the knots were also held to be variable, but each one had to contain at least two atoms of copper. The presence of such locally high concentrations of copper would cause solvent lattice distortion in their neighbourhood, and make slip more difficult by the same mechanism as was thought to cause hardening by a dispersion of precipitate particles, *i.e.* by the Jeffries and Archer slip interference theory. Merica pointed out that certain difficulties arose even in connection with the knot theory which were not substantiated by all the experimental data available. One difficulty was in connection with the results of x-ray investigations. If copper atoms did indeed segregate to preferential places in the lattice, then since copper was known to decrease progressively the lattice parameter of aluminium as it went into solid solution, the knot regions, being richer in copper, should exhibit a decreased spacing. x-Ray photographs

gave no evidence to indicate the presence in the solid solution of regions of different lattice parameter. The x-ray lines obtained were unaltered in position and quite as sharp as those existing in fully aged aluminium-copper alloy. The only change found was when $CuAl_2$ was actually precipitating during ageing at 200°C, when a diffuseness appeared in the x-ray line spectra. The x-ray evidence was thus not in complete support of the theory, but Merica pointed out that x-rays were known to be relatively insensitive for detecting local lattice distortion and so the evidence did not prove definitely that knot regions did not exist.

A further difficulty was that the diffusion of copper atoms to localized regions in the lattice to form knots would be essentially a process of uphill diffusion, that is, diffusion taking place against the concentration gradient rather than with it, as would normally be expected. No explanation of this was offered. Nowadays there would be no serious objection to this, since diffusion is now known to be controlled not by a concentration gradient but an activity gradient.

Merica also mentioned a so-called incubation period, *i.e.* the time which elapsed from the moment ageing began until the hardness showed any change. Although the hardness did not change during this time, changes in other properties such as the electrical resistance and the density took place, and it was therefore assumed that these latter changes were brought about by the diffusion of atoms to the places in the lattice where knots were about to form, but that until the knots had actually formed no changes in the hardness would be observed.

Of all the points in connection with ageing theory that remained unsolved, however, in Merica's opinion the most awkward related to the retrogression effects. He thought that these might be explained by assuming that the type of knot which formed prior to the formation of an actual precipitate particle at high temperatures was different from that at low temperatures. This would therefore mean that upon suddenly changing the temperature of ageing from room temperature to a higher temperature a period of time would have to elapse before the type of knot which applied to high temperature ageing replaced the type which was already present in the alloy and which was characteristic of the lower temperature. During this time a softening of the material could take place. Merica was thus inclining towards the view that high and low temperature ageing involved essentially different states in the solid solution.

The action of cold working, which had been found to speed up ageing, was explained simply by assuming that such working led to increasing instability in the lattice, which led to an increased rate of diffusion, so accelerating the whole process.

Merica also mentioned the effects associated with the presence in an aluminium-copper alloy of additional elements such as iron and magnesium, which could retard and accelerate the rate of ageing respectively. These effects could be brought about by the presence of only a relatively small amount of the element in question. Little explanation had been offered for such effects, although it was thought that the distortions which would be introduced into the solid solution lattice by the presence of a third element might be an important factor. Merica himself considered that such foreign elements might affect the initial knot formation by supplying more centres for the formation of knots, or alternatively might cause local distortions which would render knot formation more difficult.

From the above it can be seen that Merica largely rejected the idea of hardening by a large number of fine discrete particles and attached considerable importance to the changes prior to precipitation which have so often found their way into later theories.

Support for the knot theory was given by JENKINS and BUCKNALL,[24] who found that there was a logarithmic relationship between the time to reach the maximum on the hardness/time curve at constant temperature and the temperature itself. Thus, plotting the logarithm of the time to reach maximum hardness in an ageing alloy against the reciprocal of the absolute ageing temperature gave a straight line. A similar relationship was also found for the logarithm of the time required to reach maximum electrical resistance when plotted against the reciprocal of absolute temperature. The relationship could be expressed:

$$t = Ce^{m/T}$$

where t is the time of ageing in hours (to maximum hardness or resistance), T is the absolute temperature, and m and C are constants. This relationship was found to hold for a range of ageing systems (*Figure 4*).

It was pointed out that formation of knots in the ageing alloy would be in essence a diffusion process, as postulated by Merica. The effect of temperature upon the rate of intermetallic diffusion was known to

be logarithmic in nature, the rate increasing markedly with temperature. If the knot theory were correct, the rate of formation of knots would control the whole of the ageing process, so that it would not be surprising to find a logarithmic relationship between the time taken to reach a definite stage in the ageing process, in this particular case the point of maximum hardness, and the reciprocal of the absolute temperature. The observed relationship thus supported the theory.

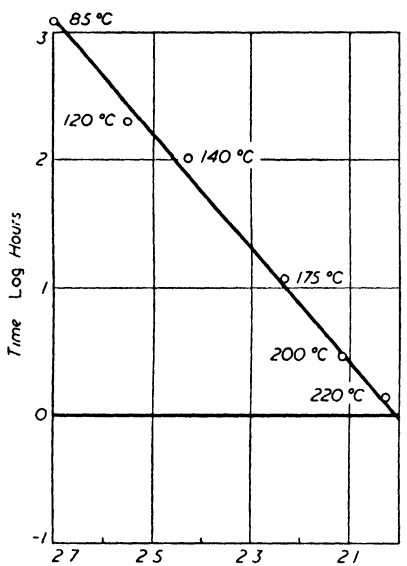


Figure 4. Attainment of maximum tensile strength during ageing of the aluminium alloy 'Lautal' (Bohner)

PORTEVIN and CHEVENARD[25] also found a logarithmic relationship to apply to the expansion of an alloy of aluminium from which Mg_2Si was precipitating, and though the form of the relationship was not exactly the same as that found by Jenkins and Bucknall, it could be reconciled. Portevin and Chevenard considered that ageing took place in two stages, the first of which they regarded as a disturbance of the homogeneity of the solid solution, due to the migration of atoms to the point where precipitation would eventually take place, and the second stage being one of true precipitation.

JENKINS and BUCKNALL[26] also discussed the actual mechanism of ageing. Their ideas were in essence very similar to the knot theory, involving the formation of localized segregate zones prior to actual precipitation. Even in those cases where hardening took place with the concurrent formation of precipitate particles it could well be that the observed hardness phenomena arose solely from changes prior to precipitation, *i.e.* the formation of the localized segregates rather than the actual formation of the precipitate. This could come about since even when some precipitate particles were visible there would probably be others present having sizes down to the smallest possible, and also localized regions where precipitation had yet to take place; these latter would lead to the hardness changes.

TWO STAGE MECHANISMS

GAYLER[27] commented upon the significance of the logarithmic relationship and stated that it seemed that the point of maximum hardness obtained on ageing at constant temperature had not a very fundamental meaning. If the theory which she and Preston had put forward were accepted, then in an ageing alloy there would be two stages, namely the rejection of atoms from solid solution with possible formation of molecules, and the coagulation of these to form definite crystalline particles which would grow to larger sizes. It was assumed that the lattice strain caused by the first stage would eventually be removed by the second, and the point of maximum hardness would therefore be the point at which the coagulation had begun to overtake the initial rejection, and would have little physical meaning. If, however, one considered electrical resistance rather than hardness, there was more significance to be attached to the maximum found on ageing at certain temperatures since it was assumed that the resistance changes were connected almost entirely with the initial stage of the ageing. Resistance measurements should therefore be more fundamental and preferable to hardness measurements which were, however, much simpler to make and therefore more frequently used.

Jenkins and Bucknall considered that the maximum hardness attained during ageing decreased as the ageing temperature increased. This was contrary to other viewpoints, in which it was considered that the maximum hardness was attained at some intermediate ageing temperature since the hardness/time curve at low temperatures was found to increase and reach an approximately constant value, less than the maximum value which could be obtained at some higher temperature, where, however, after long times the hardness decreased. Jenkins and Bucknall suggested that the reasons for these observations could be that the time required to reach maximum hardness at room temperature would be very long since, as they had shown, a logarithmic relationship connected hardness and temperature. Hence it would appear that if the experiments at low ageing temperatures had been conducted for much longer times then maximum values would in fact have been reached, which would have been higher than those reached at more elevated temperatures.

A point of considerable interest noticed during ageing was that the

various mechanical properties of a material were by no means all affected in a parallel manner. Thus although in certain cases the hardness of an alloy showed marked changes, it did not always follow that the other mechanical properties, such as tensile strength, showed similar changes. This was a point which had been discussed earlier by Rosenhain,[28] who had quoted several examples. Pure copper, for instance, containing 0·7 per cent iron, could have its hardness raised from 35·8 to 52·5 V.D.H. by quenching and ageing. In spite of this comparatively large increase in hardness, however, the tensile strength was only raised from 13 to 15 tons/sq in. With copper containing phosphorus a similar effect was observed; with 0·95 per cent phosphorus, quenching and ageing produced a change from 43 to 52 V.D.H. in the hardness, while the tensile strength showed no appreciable alteration. In copper-beryllium alloys, on the other hand, the hardness and the tensile strength both changed markedly along similar lines during ageing treatment.

Rosenhain had discussed the phenomenon at length and concluded that the reason for the observed variations was that the ageing treatment had a different effect upon the strength of the grain boundaries and the strength of the interior of the crystals. This, coupled with the fact that the type of deformation imposed in the tensile test was not the same as in the hardness test, had led to an explanation. In the light of present knowledge it seems that a very exact correlation between the results of tensile and other mechanical tests could not be expected, as in nearly all cases the types of deformation imposed in the tests are very different, and the behaviour of the alloy under test at different stages of ageing would be expected to show corresponding differences.

In 1935 Fink and Smith[29] demonstrated an important fact in connection with lattice parameter measurements during the ageing of an aluminium-magnesium alloy containing 10 per cent of magnesium. This had been solution treated at 425°C, quenched, and then aged at 100°C and 300°C for different times. On subsequent microscopical examination it was found that with an ageing temperature of 100°C precipitate was visible under the microscope at a magnification of 500 after four hours ageing but x-ray examination revealed no change in lattice parameter, nor was it possible to detect any change

even in specimens aged for as long as thirty days. At 300°C precipitate was visible after ageing for one hour, but a change in the lattice parameter did not occur until after approximately five hours (*Plate VII*). These results clearly indicated that lack of alteration in the lattice parameter of ageing alloys could not be regarded as definite evidence that precipitation was not occurring, as had been formerly assumed.

THE θ′ PHASE

A very important discovery in connection with the ageing of aluminium-copper alloys was made by WASSERMAN and WEERTS[30] in 1935. They found that the phase which first precipitated from solid solution on ageing such alloys was not the equilibrium phase of $CuAl_2$ but a similar one which, however, differed in the size of the unit cell. They did not determine the full crystal structure of the phase but found that the amounts of the equilibrium and non-equilibrium phases of $CuAl_2$ present in aged alloys varied with the ageing time, the non-equilibrium phase being gradually transformed into the equilibrium one as ageing was prolonged. At high temperatures transformation to the equilibrium phase was rapid and practically complete, whereas at lower temperatures the non-equilibrium phase was very persistent. This was a discovery of some importance as the different stability of the phases at low and high temperatures could possibly be adduced to explain the different behaviour found on ageing aluminium-copper alloys at low and high temperatures.

This was in fact done by FINK and SMITH[31] in the following year. They pointed out that the simple precipitation theory of ageing, *i.e.* the theory which attributed all the ageing phenomena to the formation of small particles of precipitate alone, required modification to take into account certain anomalous facts. However, anomalies such as the manner in which the lattice parameter, density, and electrical resistance changed only arose if it were assumed that the precipitation of very fine particles throughout an alloy led to the same effect upon its properties as the formation of much larger particles taking place under equilibrium conditions. It had already been shown that the lattice parameter could not be taken as a criterion of precipitation,[29] and now Fink and Smith dealt with the case of aluminium-copper alloys in which the presence of precipitate particles and a change in

the lattice parameter had only been found formerly when the ageing of the alloy was in such an advanced state that softening had set in.

Together with other methods, microscopical examination was used; and a special technique was developed consisting of very careful polishing followed by consecutive etching treatments in two different solutions, which had to be carried out very carefully as slight variations in the polishing or in the time of etching led to different appearances of the specimens under the microscope. The technique when carried out correctly, however, enabled evidence of precipitation to be detected in the very early stages of ageing. This evidence was not the revelation of precipitate particles but of certain regions of the crystal grains along which it was assumed that precipitation had occurred in a very fine form. These regions took the form of faint lines crossing the crystal grains, and appeared to be the traces of crystallographic planes, which were thought to be the (111) set (*Plates VIII* to *XI*). Evidences of precipitation could be observed in the alloy after only very short ageing times, but lattice parameter measurements showed no change until much later. Furthermore, when the lattice parameter did begin to change it did not do so uniformly throughout the whole of the material, as only a small percentage underwent a change while the bulk retained the original parameter value. The amount which underwent a change in parameter increased as the time of ageing was increased, but some of the material of the original parameter still remained even after relatively long periods of time. x-Ray analysis was used in order to check the nature of the precipitating phase, and $CuAl_2$ and the non-equilibrium phase were found to be present in amounts depending upon time and temperature of ageing. ($CuAl_2$ was called θ by Fink and Smith, and the non-equilibrium phase θ'. Later workers have used α $CuAl_2$ and β $CuAl_2$ to denote the non-equilibrium and equilibrium phases respectively.)

The fact that the amounts of the two phases varied with time led Fink and Smith to put forward an explanation for the changes in density which occurred on ageing. It was known that the observed changes in density were not in agreement with those to be expected if $CuAl_2$ were merely being precipitated. The density of a solid solution of 5 per cent copper in aluminium being 2·803, and that of $CuAl_2$ 4·35, the aged alloy should have had a density of 2·796, but the actual

density observed after ageing at 220°C was less. On the other hand, if the density of θ′ was taken as 3·99, and if this was formed upon ageing rather than $CuAl_2$, then the density of the alloy would become 2·787, which would be much closer to the actual experimental figures though slightly low. An exact agreement could, however, be obtained by assuming that the alloy contained a mixture of the two phases or that the precipitation of θ′ was not complete. The fact that the density changes observed were dependent upon the temperature at which the ageing was carried out would be explained by the more ready transformation of the alloy to the equilibrium phase at the higher temperatures, so that the actual change in density at the higher temperatures would be less than the change at low temperatures, a deduction in agreement with the observed facts. It would also be expected that a minimum density would be reached during ageing, since this would represent the point at which the transformation of θ′ to $CuAl_2$ was taking place in considerable amount.

The results obtained from metallographic examination and tensile tests indicated that in all cases precipitation occurred before any substantial changes took place in the values of yield strength and elongation (*Figure 5*). It seemed highly probable that the actual precipitation of particles could account for the changes in these properties, and that there was therefore no need to complicate the theory of ageing by postulating that pre-precipitation changes such as the formation of knots were responsible for ageing phenomena.

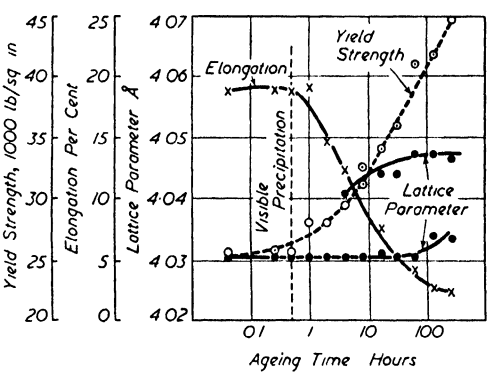


Figure 5. Time for visible precipitation and change in lattice parameter, yield strength, and elongation on ageing an aluminium-copper (5·17 per cent Cu*) alloy at 160°C (Fink and Smith)*

The behaviour of the electrical resistance of the alloy, which was one of the outstanding anomalies, was attributed to the fineness of the particles formed at lower ageing temperatures, for according to Fink and Smith the behaviour of the fine particles in their effect upon the properties of the alloy might not be the same as that of larger particles.

Thus at low temperatures, where the particles were very fine to begin with, the resistance might increase initially, whereas at higher temperatures, where coarser particles would form, it might decrease from the start. This would then be in agreement with the observed facts.

Fink and Smith drew attention to the fact that in the aluminium-copper system ageing appeared to reach completion in certain parts of the crystal grains before it had begun in other portions. This followed from their observations concerning the way in which the lattice parameter of the alloy changed. NORTON pointed out that there appeared to be two distinct types of ageing systems, namely those which behaved as outlined above and those in which the precipitation appeared to take place throughout the entire crystal structure at the same time and rate. Prior to Fink and Smith's work it had been thought that the aluminium-copper alloys belonged to the class showing uniform precipitation, but it seemed that either the earlier view was incorrect or that the same system could under certain conditions exhibit two different behaviours.

LOCALIZED PRECIPITATION AND LATTICE DISTORTION

The localization of precipitate upon certain crystallographic planes in the material could have two explanations. Either the alloy had an inherent tendency for such localization, or alternatively, for some reason connected with the treatment of the alloy, certain places became more suitable for precipitation. Fink and Smith mentioned that the planes upon which preferential precipitation appeared to take place were the (111) planes, and RHINES pointed out that these were the planes upon which it was known that slip took place when an aluminium-copper alloy was deformed. Thus it was possible that the treatment of the alloy prior to ageing so as to cause plastic deformation to take place caused certain regions of the crystals, that is the slip planes, to suffer plastic deformation. Since it was known that the rate of ageing of alloys was in general increased by plastic deformation, Fink and Smith's microscopical observations could be due to the fact that deformation had occurred on the slip planes with consequent acceleration of the ageing at such places. Though the work of Fink and Smith was important from several points of view, it did not settle which changes in an ageing alloy were responsible for the observed effects. The fact that it had been shown that precipitation could take place

without change in the lattice parameter did not affect the knot theory, for it was assumed by most supporters of the theory that the formation of the knots would lead to no detectable change in the parameter values of the alloy. Fink and Smith argued however, that since the presence of precipitate particles could be shown in an ageing alloy before any appreciable change could be observed in the properties, then any pre-precipitation processes such as the formation of knots could not exert any very large effect upon the properties of the material, and should therefore be ignored in explaining ageing changes.

Since precipitation appeared to be highly localized in the alloy, it would seem that there was insufficient evidence to indicate that it would alone suffice to explain the changes in properties of the alloy as a whole, and it would have seemed more reasonable to suppose that though the localized precipitate could cause the changes, it was possible that pre-precipitation changes taking place in the large bulk of the material were actually responsible. Thus knots could have been forming inside the crystal grains at those points which had not suffered any deformation, and such knots could have been responsible for the ageing changes. At the points where distortion had taken place in the crystals knot formation would take place much more rapidly and pass into the stage of the formation of minute precipitate particles. No effect of the initial formation of these knots would have been felt upon the properties of the material, however, due to their localized nature. It is reasonable to assume that local stresses were in fact set up on the (111) planes in Fink and Smith's work, since the act of quenching the material from the heat treating temperature into cold water would be sufficient to have caused plastic deformation. Later this was shown to be so.

The following year GAYLER[32] put forward a new theory which involved pre-precipitation ideas. In an earlier paper Gayler and Preston had concluded that ageing was essentially a two-stage process, both parts of which involved the formation of precipitate particles, although the size of the particles varied in each. Gayler's later theory retained some of the previous ideas but differed in several important respects. Ageing was again considered a two-stage process, but in the first stage the diffusion of solute atoms was supposed to take place in such a way that they concentrated upon certain crystallographic planes in the solid solution, the planes on and about which precipitation

would eventually take place. This resulted in a distortion of the lattice, which in turn resulted in an increase in the hardness and the electrical resistance, but no change in lattice parameter was supposed to occur. In the second stage, which occurred either immediately after the first or occasionally almost concurrently with it, molecules were formed by the interaction of the diffused solute atoms with either solvent atoms or sometimes, and depending upon the actual precipitating phase, with other solute atoms. This formation of molecules would proceed and at the same time there would be an agglomeration of the molecules to form small particles of precipitate. The agglomeration would result in a decrease in the strain in the lattice, and consequently also a decrease in the electrical resistance and in the rate of hardening. The growth of the small precipitate particles however, would generate localized stresses which would cause diffuseness in the x-ray spectra. Up to this point in the ageing it was considered that the particles were still on the lattice of the solid solution, but eventually, as they grew in size, they would come to the point at which the lattice was no longer able to withstand the stresses set up and consequently the particles would break away from the lattice and precipitation proper would have begun. The point of actual precipitation would be accompanied by the release of the stresses which had been set up as the particles grew.

In the new theory, Gayler thus explained ageing as being due both to the actual precipitation of fine particles of almost molecular dimensions, as had been postulated in the earlier theory, and also to the changes in the lattice which took place prior to precipitation proper. The two stages were assumed to take place more or less together, and their relative amounts were considered to be governed by the ageing temperature and time.

TEMPERATURE CONSIDERATIONS

It was considered that for all ageing systems there existed a range of temperature over which the changes characteristic of ageing could be made to take place. This temperature range was thought to be governed by the character of the solid solution at the lower temperatures, and by the solid solubility line at the higher temperatures. This would mean that for some alloys, such as the aluminium-copper, ageing effects could be produced at around room temperature, whereas for others,

such as copper-beryllium, they could only be produced at higher temperatures. The two stages of ageing would take place in varying amounts at different temperatures within the range, and depending upon their relative predominance at any temperature the behaviour of the alloy would differ (*Figure 6*). At low temperatures the ageing would proceed very slowly, so that the second stage would not occur for a very long time, if at all. The behaviour of the alloy at such temperatures would thus be governed almost entirely by the first stage. At higher temperatures, however, the first stage would take place rapidly, and the overall behaviour would be controlled by the changes characteristic of the second stage. It was further assumed that both stages would give rise to changes in the properties of the alloy, and that those characteristic of the second stage would gradually replace those characteristic of the first as ageing proceeded. Thus the properties of an alloy, as a function of ageing time at constant ageing temperature, could be represented by the sum of two curves, one of which would represent property variations due to the first stage, and the other variations due to the second stage. A series of schematic curves could then be drawn to indicate the overall property changes which could be expected in an ageing alloy at various temperatures within its ageing range. At high temperatures both stages would be rapid, so that for instance after an initial rapid increase in hardness there would quickly be a coagulation of the precipitate, resulting in a rapid softening. At very low temperatures only the first stage would be effective, taking place only slowly, so that there would be a gradual change of the properties with time. At intermediate temperatures the property changes due to the first and second stages would succeed each other. The point at which the second stage began to become

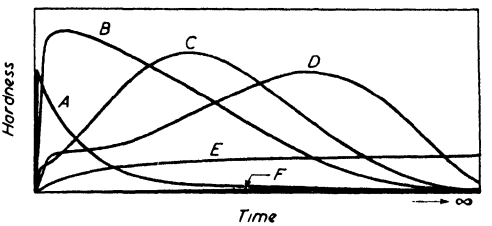


Figure 6. Curves representing time/hardness changes at different ageing temperatures. A *Upper temperature limit of ageing. Both stages of ageing take place excessively quickly.* B *Second stage of ageing, perhaps slight indication of first stage.* C *and* D *First and second stage of ageing both apparent.* E *First stage of ageing and perhaps entry of second stage at an excessively slow rate.* F *Lower temperature limit of ageing. Both stages of ageing may take place excessively slowly (Gayler)*

appreciable would occur sooner the higher the temperature, but if it did not take place until after the changes due to the first stage had passed their most effective point, there could be a fall in the property/time curve before it rose again due to the onset of the changes due to the second stage. Thus there would be a maximum value for a property such as hardness, both for the first stage and the second stage, the actual values being dependent upon the temperature (*Figure 7*). The curves obtained, showing the variation of properties as a function of time at various temperatures, would differ in shape due to relative influences of the two effects at different temperatures. Some curves would show only one maximum of a fairly sharp nature, whereas others could show two maxima, joined by a rather flat portion.

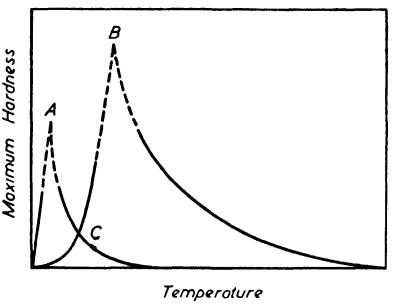


Figure 7. Relation between maximum hardness and temperature (Gayler)

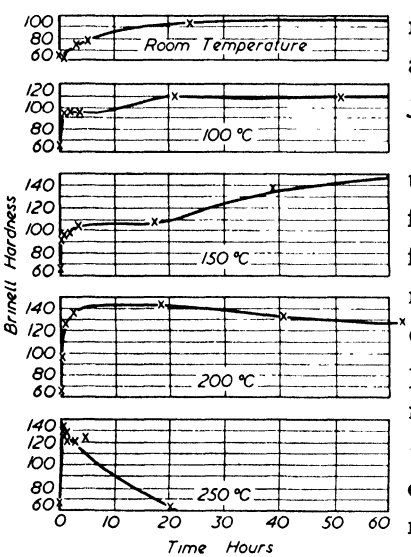


Figure 8. Aluminium-copper alloy heat treated at temperature indicated after quenching from 550°C (Gayler)

This theoretical consideration of the possible types of curve resulting from different ageing treatments was found to be in agreement with experiment. For duralumin, experimental curves showed evidence of the two processes, each one of which gave a maximum in the property/time curves (*Figure 8*). The first stage of ageing caused an increase in hardness, which reached maximum value at temperatures slightly above room temperature. It was not known whether the second stage of ageing occurred at room temperature or not, but it was thought that this could be checked by using a method of measurement known to be very sensitive to internal changes taking place in a metallic lattice. Electrical resistance measurements were thought to be satisfactory.

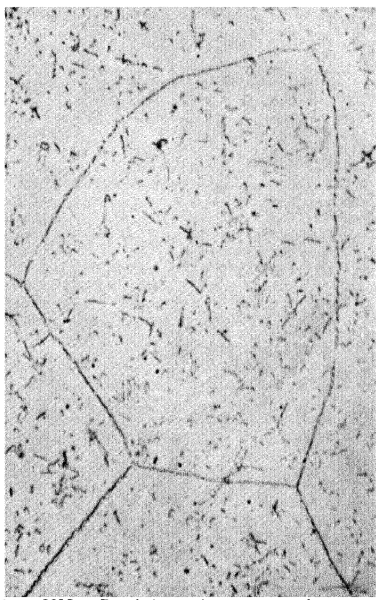


Plate VII. Precipitate in a cast aluminium magnesium alloy (10 per cent Mg*) heated eighteen hours at 425°C, quenched in water and aged four hours at 300°C. Etched five seconds in 0·5 per cent* HF *solution (× 500) (Fink and Smith)*

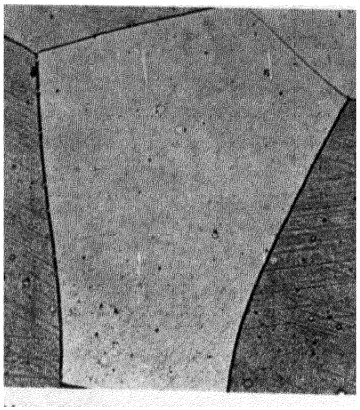


Plate VIII. Aged 3 months at room temperature (× *100*)

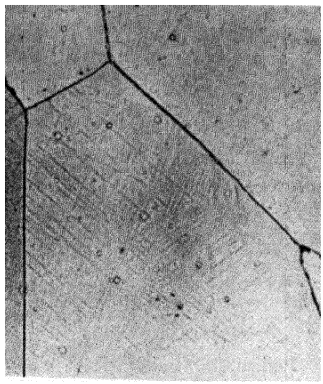


Plate IX. Aged ½ *hour at 160°C* (× *100*)

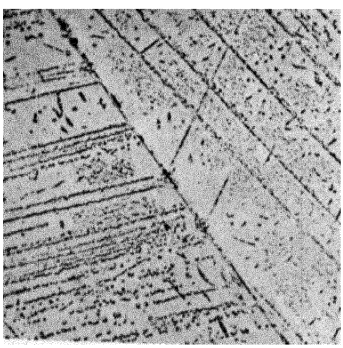


Plate X. *Aged* 2¼ *hours at 200 C* (× *500*)

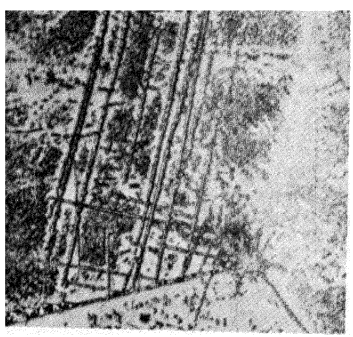


Plate XI. Aged 256 hours at 160 C (× *500*)

Plates VIII to XI. All aluminium-copper alloy (*5·17 per cent* Cu) *heated 20 hours at 540°C, quenched in water and aged. Etched in 25 per cent* HNO_3 *at 70°C, washed, and then etched in a solution containing* NaF, HNO_3, HCl *and* H_2O (*Fink and Smith*)

The resistance of an alloy containing 4 per cent copper, 0·5 per cent magnesium, 0·3 per cent silicon, and 0·11 per cent iron showed on ageing at room temperatures an initial increase and then remained constant. Whilst other results showed that with an alloy containing 4 per cent copper and 0·5 per cent magnesium the resistance remained approximately constant over a period of from three to eight days and then gradually decreased. This seemed to indicate that the second stage of the ageing was eventually making its appearance and overlapping the first. These and other results therefore indicated that at temperatures in the region of room temperature most of the ageing changes were probably due to the first rather than the second stage.

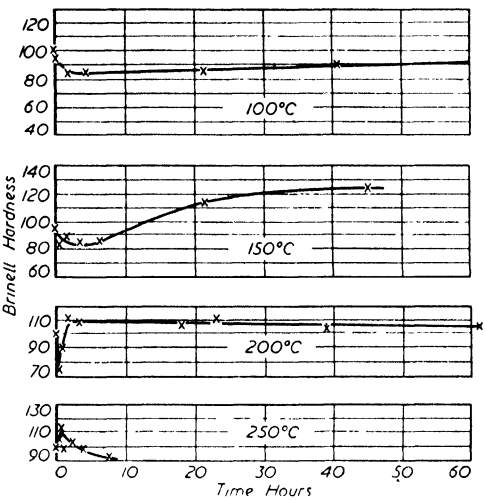


Figure 9. Aluminium-copper alloy age hardened at room temperature and then heat treated at various temperatures (Gayler)

In connection with retrogression effects (*Figure 9*) Gayler quoted some of the results which had been obtained by MASING and KOCH,[33] who had shown that the initial decrease in hardness which took place on the second ageing was always associated with a pronounced decrease in electrical resistance. Masing and Koch had sought to explain these effects using the theory of VOLMER and WEBER,[34] which had been put forward in connection with crystallization and nucleation. The theory indicated that for a given temperature there was a minimum size for stable crystal nuclei, and that this minimum size decreased with temperature. Initially, on ageing at room temperature, small nuclei would form and result in a change in properties. If the ageing temperature was then raised, the smallest particles would be unstable and dissolve, a process which would result in a decrease in hardness. On continued ageing at the higher temperatures, however, more nuclei of the size stable at the new temperature would form, and these would cause the hardness to increase again before it eventually decreased due to extensive coagulation.

Gayler, however, offered an alternative explanation based upon the extent to which the stages of ageing proceeded at low and high temperatures. It was postulated that raising the temperature after ageing at low temperatures would first increase the rate at which the second stage proceeded. This initial rapid increase would cause a decrease in hardness and at the same time a decrease in the resistance. As the second stage became more pronounced the stresses in the lattice would be increased again, which would lead to the observed subsequent increase in hardness.

WAVE MECHANICAL INTERPRETATION

MOTT[35] introduced modern wave mechanical ideas in discussing Gayler's theory and pointed out that, according to wave mechanics, the resistance of a solid solution was regarded as being due to the wave associated with the electrons being scattered and diffracted by the solid solution lattice, and therefore unable to penetrate the lattice at all easily.

If one considered the distribution of copper atoms in a solid solution of copper in aluminium, the maximum resistance would occur when the atoms were arranged in such a manner that they exerted the maximum possible scattering effect upon the electron wave. This would occur when the size of the points scattering the waves were of the order of a wavelength. Assuming a wavelength of about four to five interatomic distances, maximum resistance would occur when any copper groupings occurred of approximately such a size. Maximum resistance would therefore occur only after there had been some segregation of the atoms, and the maximum resistance would not be expected in a uniform solid solution as previously assumed. According to Mott, it was possible that the segregates in the aluminium-copper alloy responsible for the scattering of the electron waves were the same segregates postulated by Gayler in the first stage of her ageing mechanism, though this could not be taken as proved, since strains present in the lattice could possibly exert a considerable effect.

Several objections[36] were put forward to the idea of the segregation of solute atoms to specified planes, on the grounds that such a process would be akin to the formation of a superlattice which, it was well known, resulted in a decrease and not an increase in the electrical resistance. The parallel did not, however, seem to be very exact, and

in view of Mott's conclusions it was not a very serious criticism of the theory.

FURTHER THEORETICAL DEVELOPMENTS

It should be noted that Gayler did not attribute the effect of the initial stages of ageing upon the mechanical properties to the same mechanism as Merica in his knot theory, that is, to the slip blocking action of the localized segregate regions in the lattice. Gayler thought the hardening was due entirely to the state of strain set up in the lattice by the growing segregate regions. A type of strain hardening would result which would give an increased resistance to slip.

During the discussion of Gayler's paper, others suggested mechanisms of age hardening some of which are now of considerable interest in view of Rohner's recent theory of ageing. It was suggested[37] that in the initial stages of ageing of aluminium-copper alloys the aluminium lattice would be in a highly disorganized condition and perhaps contain imperfections of a localized nature. There was a possibility that such imperfections were the cause of the high yield point, and possibly electrical resistance, obtained by ageing since Taylor had shown that a lattice possessing a large enough number of imperfections (dislocations) was stronger than one which contained a lesser number. This could therefore correspond to the first stage of ageing and the second stage would be due to the formation of actual precipitate particles perhaps of a critical size.

Another idea[38] related to the relative sizes of the aluminium and copper atoms and the fact that dislocations were known to be regions in the lattice where the spacing was abnormal due to solvent atoms being more tightly packed together than in a perfect lattice. It was suggested that the copper atoms, being smaller than the aluminium atoms, might diffuse so that they went into those regions of the lattice which exhibited this smaller spacing. This would produce a better fit in the lattice, which might therefore result in an increase in strength. These theories were thus tending to the idea that the making or repairing of defects in the matrix lattice produced by the movement of copper atoms prior to the formation of precipitate particles was the cause of the increased strength during ageing.

COHEN,[39] investigating the ageing of silver-copper alloys, arrived at conclusions similar to Gayler's. Before Cohen's work the silver-copper

alloys were thought to be clear cases of alloys which hardened by the formation of distinct precipitate particles. Cohen, however, deduced that there was another possible mechanism of ageing, which he identified with knot formation. It appeared that the mechanism changed at different temperatures, a fact which was explained by assuming that the rates at which knot formation and actual precipitation took place were different at different temperatures. He suggested, therefore, that all the other systems which were thought to be definite cases of hardening by means of precipitate particles might also harden by the same pre-precipitation mechanism which had now been found to occur in the silver-copper alloys. Definite evidence was found in the hardness/time curves of the silver-copper alloys for two stages in the ageing, since the curves showed two maxima, the first of which was attributable to knot formation and the second to the formation and growth of precipitate.

It was pointed out that even if it were assumed that these two stages were present in all ageing alloys it did not necessarily follow that they would be easy to detect. It might happen that the nature of the material was such that the pre-precipitation stage took place very rapidly, so that the net ageing changes would be dependent almost entirely upon the second precipitation stage. In such a case it would be extremely difficult to show that ageing could take place without precipitation unless the correct low temperature were found.

Gayler and Cohen regarded the existence of two maximum values in the hardness/time curves as being in no way extraordinary, but rather to be expected on the two-stage mechanism postulated. Many investigators were, however, of the opinion that the curves which showed two maxima were due to faulty experimental technique, probably connected with inhomogeneous material, and this view was considerably strengthened in 1938 when FINK and SMITH[40] put forward a most ingenious explanation of such effects and confirmed their ideas experimentally. Their explanation was based upon the assumption that plastic deformation could accelerate ageing, an assumption which was borne out by the experimental results of previous workers. It was further suggested that plastic deformation could originate in specimens purely from the act of quenching, and experiments showed that this was so. Polished specimens were subjected to solution heat treatment followed by quenching, and were then examined before repolishing

(*Plate XII*). The grain boundaries and slip planes were clearly defined, indicating that plastic deformation had occurred and had produced slight relief effects. The specimens were then repolished after ageing at 200°C, and on etching, precipitate was found in just those regions which had suffered plastic deformation during the quenching (*Plate XIII*). Specimens quenched in boiling water, or in oil at 153°C, showed no slip lines or grain boundary relief due to the quenching (*Plates XIV, XV*), and on ageing such specimens the precipitation was general over the surface of the crystal grains rather than in localized regions.

Plots of tensile and yield strength curves as a function of ageing time showed the presence of an initial peak, attained after a relatively short time, followed by a decrease and then a further peak.

The results were explained by assuming different rates of ageing of the deformed and undeformed regions of the crystal grains. The deformed portions would suffer a more rapid rate of ageing, so that the overall behaviour of the alloy would be represented by the sum of two curves, giving the variation of properties with time in the deformed and undeformed regions respectively. Ageing would proceed rapidly in the deformed regions, so that they could age past their maximum before a maximum was reached in the undeformed regions. There would thus be a decrease in properties after a certain ageing time, representing the point at which the softening of the deformed regions was taking place before the undeformed regions had reached their maximum. The magnitude of the first ageing peak would depend on the amount of plastic deformation to which the alloy had been subjected.

Experiments on water quenched specimens showed that two peaks could be distinguished on the ageing curves, but with specimens quenched into oil at 153°C two peaks were not observed and the hardness was found to change continuously in a manner similar to the second part of the curves obtained for water quenched specimens. An anomalous effect was found in specimens quenched and aged in oil at 153°C, which did not show hardening until after approximately one hour, while specimens quenched and aged in oil at 100°C started to harden almost immediately. Since a higher ageing temperature should cause more rapid hardening, the specimens aged at 153°C should have aged more rapidly. An explanation offered was that the

specimens quenched in oil at 100°C received a greater amount of plastic deformation than those quenched in oil at 153°C. This larger amount of plastic deformation was sufficient to cause a more rapid ageing despite the lower ageing temperature.

Fink and Smith thus explained the double ageing peaks, which were an integral part of Gayler and Cohen's theory, as being an experimental effect, and not in any way related to the fundamental mechanism of ageing. The appearance of precipitate in localized regions of grains had of course been observed earlier, but the effect upon the properties of the alloy had not been considered.

COHEN[41] pointed out that if plastic deformation was significant, certain portions of the specimens might over age before the bulk of the material reached maximum hardness. The net maximum hardness attained by such specimens would not be as great as the hardness which would be attained when the ageing rate was uniform. It was to be expected, therefore, that the rate of quenching would exert a significant effect upon the maximum hardness attained by an alloy. Fink and Smith's data showed no such effect; this would therefore seem a criticism of the explanation they offered. Neither could accelerated precipitation explain the contraction which was known to take place when aluminium-copper alloys were aged at low temperatures. Precipitation of any kind, whether of a transition phase or the equilibrium phase, would result in an expansion. WASSERMANN[42] questioned whether the lines produced in Fink and Smith's specimens were slip bands and suggested that they might be twins formed during quenching in the manner suggested by Schmid. It would appear, however, that they were in fact slip bands, as Fink and Smith found no evidence of their presence after repolishing. In reply to discussion, they expanded considerably their ideas on ageing, and gave a more detailed picture of the processes which they believed to be involved.[40]

They assumed that a random arrangement of atoms would exist at high temperatures in a solid solution but that there would be a variation with time in the concentration of solute atoms at any point. Some points might have locally high concentrations, but such arrangements would not be stable, and they would be removed by thermal agitation and diffusion. On quenching, the high temperature structure would be retained, and for aluminium-copper alloys, for instance, any localized high concentrations of copper having structures similar to the

arrangement in the non-equilibrium or equilibrium structure, which should precipitate, would be stable. On probability grounds, the precipitate formed would be the smallest stable particles at the ageing temperature, and this size would decrease with decreasing ageing temperature. This theory was similar to that developed by Einstein for the formation of a disperse phase in a colloid system, but in ageing systems a modification had to be introduced to take account of the fact that the substances taking part were crystalline, and had therefore to obey the rules relating to the formation of precipitate particles, elucidated by Mehl and his co-workers in their work on Widmanstatten mechanisms. Growth of the θ or θ' phases would take place by the diffusion of atoms up to the interface between the precipitate and the solid solution, and these atoms would subsequently become attached to the lattice of the precipitate. Such diffusion would be of the normal downhill type, since the atoms would be removed from the solid solution in contact with the precipitate, and thus leave the adjacent regions of the lattice clear for further diffusion to occur. In the early stages of ageing it was assumed that there would be only a few small localized precipitate particles, which would have little effect upon the properties of the material. As the particles grew, the point at which they began to exert an appreciable effect on the properties of the material would depend upon the system concerned. For example, in the aluminium-magnesium system it was found that in some parts of the grains the particles grew to sizes which were clearly resolvable under the microscope before any change in the yield point or the lattice parameter could be measured.

The actual changes in properties on ageing were then considered. Changes in resistance were explained by reference to wave mechanical considerations similar to those mentioned by MOTT[35] in connection with Gayler's paper, but instead of postulating that the diffracting regions were local segregations within the matrix lattice it was assumed that they were actual particles of precipitate, the initial size of which would vary with the temperature of ageing. A critical size was assumed to give maximum resistance, so that depending upon whether the initial size was less or greater than the critical, the resistance would increase or decrease at the beginning of ageing.

The changes in density which took place on ageing were regarded as due to two factors: the actual densities of the precipitate particles,

and the way in which the lattices of the precipitate and solid solution fitted together at the interface. In the initial stages the particles would be very small and the register between the two lattices would be good so that there would be negligible change in the density of the alloy as a whole, but as the particles grew, the increasing quantity of precipitate would begin to exert an effect upon the density due to increasing lack of register. Both factors would tend to cause a decrease in the density in aluminium-copper alloys. There was no explanation, however, of the initial increase in density known to occur during the room temperature ageing of aluminium-copper alloys, other than a suggestion that it might be due to impurity.

Fink and Smith considered that hardening was due not only to those particles visible under the microscope. They postulated that the maximum hardening effect was caused by a critical size of dispersion of colloidal dimensions, a proposal somewhat similar in manner to the original proposals of Merica, Waltenberg, and Scott. The important factor governing the extent of the hardening was thought to be the total number of particles present within a certain range of sizes; the total amount of phase precipitated as such was unimportant. Alloys of aluminium and nickel containing only 0·03 per cent nickel, and of aluminium and beryllium containing only 0·07 per cent beryllium, were known to show the same relative increase in hardness as aluminium-copper alloys containing 4 to 5 per cent copper. Thus the quantity of precipitate did not appear to be the most important criterion of the extent of hardening. Retrogression effects could be explained by the fact that the smallest particles stable at room temperature would not be stable at higher temperatures, so that the first effect of raising the temperature would be to destroy the smallest particles, a process which would result in softening. Growth to larger size at the higher temperature would then give a subsequent increase in the hardness. Differences in corrosion behaviour of aluminium-copper alloys after different ageing treatments could also be explained by depletion of copper in the solid solution due to precipitation. The depleted regions would have a different solution potential from those which had not undergone any depletion. At low temperatures of ageing the depleted regions could be much smaller than the distance between the precipitate particles; thus there would be no continuous depleted layer. At higher temperatures, however, extension of the depleted layer would take

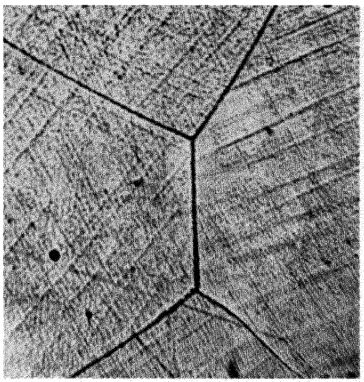


Plate XII. Polished, as-quenched in water at 25°C

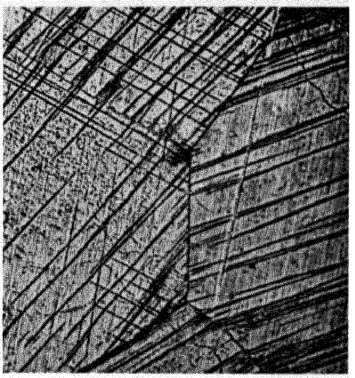


Plate XIII. Same as Plate XII after being aged 2 hours at 200°C, lightly polished and etched

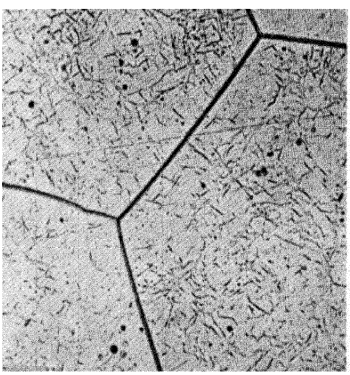


***Plate XIV.** Quenched in boiling water, aged 2 hours at 200°C, polished and etched*

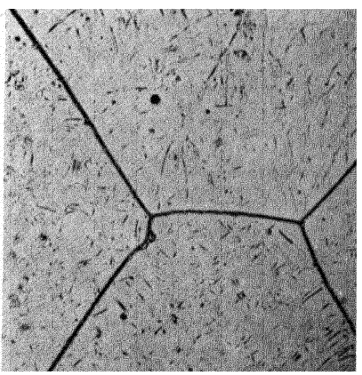


Plate XV. Quenched in oil at 153°C, aged 2 hours at 200°C, polished and etched

Plates XII to XV. Aluminium-copper alloy (5·17 per cent Cu*) heated 16 hours at 540°C (× 100) (Fink and Smith)*

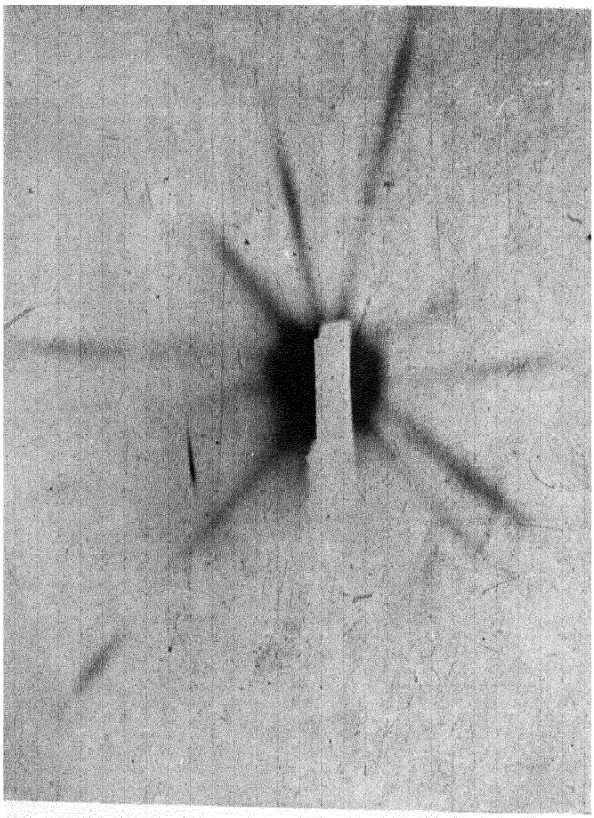


Plate XVI. ***The x-ray diagram of an aluminium-copper alloy aged at 100°C for 16 days and showing long and well defined streaks*** *(Calvet, Jacquet and Guinier)*

place due to growth of the particles. In general, growth would be most rapid at the grain boundaries, giving rise to a continuous depleted layer which would result in intergranular corrosion. On still further ageing, the growth of the precipitate particles at the boundaries and the appearance of particles in the interior of the grains would result in a uniformly depleted solid solution throughout the whole of the material. There would then be no tendency for attack at the boundaries. It was thus explained why serious intergranular corrosion arose only after ageing at certain temperatures.

Fink and Smith did not deny that there would be some stage preliminary to actual precipitation. They thought that such a stage would take the form of internal segregation of solute atoms. Their main argument against the knot or allied theories was that such theories had been put forward originally only as an explanation of the changes which took place in the properties of alloys before any visible precipitate appeared. If the precipitate could, however, be revealed before the properties changed, there seemed no reason for the retention of the idea that pre-precipitation changes produced the changes of properties; the simplest theory would be that the changes were due to particles of precipitate. According to Fink and Smith, no evidence showed that any pre-precipitation process was exerting an effect.

COHEN,[43] however, had pointed out that attempts to distinguish between the precipitation and the knot theories could only be made in those cases where ageing took place uniformly throughout an alloy, for if precipitation was localized, knot formation or allied effects taking place in those regions where precipitation was not visible might be masked. In Fink and Smith's work the ageing was not uniform, so that according to Cohen it was not possible to deduce any definite conclusions about the relative importance of the pre-precipitation and precipitation stages.

THREE STAGES OF AGEING IN DURALUMIN

The question as to the fundamental nature or otherwise of double ageing peaks was one which received further attention from Cohen in an investigation of the ageing of duralumin in 1939.[44] Measurements of hardness, electrical resistance, dilation, and lattice parameter gave evidence of three stages in the ageing of duralumin. This was observed

only in specimens aged within the temperature range 150° to 200°C (*Figure 10*). At lower temperatures, 65° to 100°C, only the first two stages were detected, and at still lower temperatures, 30°C, only the first stage.

The different stages appeared to obey the original logarithmic law discovered by Jenkins and Bucknall. The initial hardening was found to be accompanied by a large increase in the electrical resistance, the increase being greater at lower temperatures. Dilation

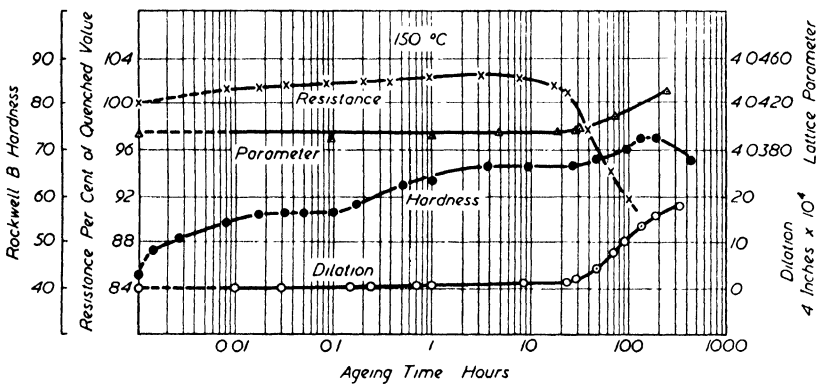


Figure 10. Ageing curves of duralumin alloy at 150°C (Cohen)

curves showed at low temperatures an expansion followed by a contraction, whereas at higher temperatures the initial expansion disappeared, and an initial contraction was followed by an expansion. At still higher temperatures the curves showed an expansion from the start. These results seemed, at first sight, to be somewhat anomalous. The initial expansion at the lowest ageing temperatures could not be due to the formation of $CuAl_2$, as it was considered most unlikely that it would form at such low temperatures. Furthermore, pure aluminium-copper alloys were known to show a contraction on ageing at room temperature. To check the nature of the low temperature expansion Cohen overaged specimens at 191°C, heated them rapidly to the solution treatment temperature, and quenched them immediately, giving no time for any solution of precipitate to take place. The specimens were then aged at room temperature in the same manner as those which had been quenched after the normal length of solution treatment. The former specimens showed little increase in hardness on ageing, since a negligible amount of solute had been taken up before quenching.

Nevertheless an expansion took place upon ageing; no contraction occurred as with specimens subjected to full solution treatment prior to ageing. It seemed, therefore, that the initial expansion was not connected with the hardening of the alloy but was due to the relief of quenching strains. The contraction which normally followed the expansion appeared to be intimately connected with hardening, and it seemed highly probable that it was connected with the same effect that produced a contraction in the ageing at room temperature of pure aluminium-copper alloys. Cohen believed that the pre-precipitation stage was followed by actual precipitation of a non-uniform nature, beginning in certain localized regions which might be either the grain boundaries or the slip planes. In duralumin it appeared that only the grain boundaries were concerned, as no evidence of slip plane precipitation was found. This localized precipitation was thought to be responsible for the changes in the alloy following the changes due to the pre-precipitation stage. Thus the resistance decreased and the alloy expanded slightly, but no change in lattice parameter was found. The slight expansion came about as a result of the slight amount of precipitation balancing the effect of nucleus formation which would take place over the majority of the grains.

The third stage of ageing was marked by the occurrence of general precipitation and was accompanied by a large expansion, a further decrease in the electrical resistance, and a change in the lattice parameter.

No change in the microstructure was observed during the first stage of hardening; a darkening of the grain boundaries occurred during the second stage; and in the third, a general darkening of the whole microstructure took place.

Cohen suggested that these three stages represented behaviour general in all ageing systems. Individual systems might differ as a result of the suppression of one or more of the stages, depending on the nature of the alloy under investigation and the conditions under which ageing was carried out. In certain cases localized precipitation could proceed so rapidly that all traces of the pre-precipitation stage would be obliterated, when the overall behaviour of the alloy would comprise two stages: first localized and then general precipitation.

GAYLER,[45] discussing Cohen's ageing peaks, suggested that the first two might be due to two pre-precipitation processes occurring at

different rates. In a complex alloy such as duralumin the precipitation of several phases might occur and the segregation of the two types of solute atom concerned in certain crystallographic planes in the alloy might produce two separate hardening peaks. Cohen regarded this as unlikely, for it did not appear that the amounts of impurity present in the material would be large enough to affect the hardness in the way observed.

Cohen emphasized that it might be very difficult to detect all three stages of ageing in some alloys, as the alloys might be very sensitive to variables such as the presence of slight strains, which might be sufficient to accelerate the precipitation at certain points. This seemed to be so with aluminium-copper alloys, and in such circumstances the initial pre-precipitation effects would be masked, and the behaviour would appear to be due entirely to precipitation.

The initial expansion found by Cohen on ageing at low temperatures was not connected with any internal changes of solute atoms prior to precipitation, and the correct sequence of volume changes connected with ageing at room temperature appeared to be an initial contraction followed by an expansion. At higher temperatures the initial contraction became less marked, and disappeared at higher temperatures still.

In spite of the evidence presented by Cohen's and Gayler's work, the double and treble ageing peaks were still attributed by many to experimental technique. The peaks were difficult to reproduce, indicating that they might be controlled by external variables.

THE GUINIER-PRESTON ZONES

Evidence was obtained in 1938[46] and 1939, however, to show that in the ageing of aluminium-copper alloys there was a definite segregation of copper atoms to certain regions of the matrix lattice prior to the actual formation of precipitate particles of θ' and $CuAl_2$. This was provided by Guinier and Preston[57] and then again by Calvet, Jacquet, and Guinier, and was an important step forward in elucidating the method by which precipitate particles formed.

Calvet, Jacquet, and Guinier investigated the ageing of a high purity aluminium-copper alloy by metallographic and x-ray means. In their paper they summarized previous results obtained on aluminium-copper alloys by these techniques. They concluded that considerable doubt was to be attached to some of the x-ray work.

Schmid's and Wassermann's claims to have shown actual precipitation of $CuAl_2$ during ageing at 150°C and Hengstenberg's and Wassermann's[23] indications of internal movement of copper atoms were thought to be not above criticism. The weight of evidence showed that it had not been demonstrated with certainty that any migration of copper atoms, or precipitation, took place upon ageing before the mechanical changes had reached an advanced state. Most metallographic examinations also showed that no precipitation was visible until hardening was well advanced. About the only case to the contrary was that described by Fink and Smith, but this was regarded as difficult to accept because of discrepancies between the microscopical changes and the yield point changes, the latter lagging considerably behind the former. Fink and Smith's parameter and density measurements upon θ′ were also regarded as being open to criticism.

Calvet, Jacquet, and Guinier used electrolytic polishing in order to obtain as good metallographic results as possible, and, in their x-ray work, a very careful technique. Initial tests were made to determine the rate of solution of the copper rich phase at the solution heat treatment temperature, by microscopic analysis of material treated for different lengths of time. Since precipitate was not visible after 16 hours treatment, it was taken that 24 hours would suffice to ensure that the alloys were homogeneous. In interpreting the appearances of their specimens under the microscope, only those precipitates which could be seen under a magnification of × 1,850 when the specimen had been electrolytically polished were regarded as being present in the alloy. The x-ray transmission photographs of certain specimens showed a series of streaks (*Plate XVI*), which varied in size and intensity with the temperature and time of ageing and were correlated with internal movement in the solid solution prior to precipitation, as will be mentioned later.

A specimen was first aged at 20°C and the hardness found to increase over a period of 340 days. During this time no evidence of precipitation was detectable either within the grains or at the boundaries. Some x-ray photographs were taken of specimens after ageing for varying lengths of time at room temperature. During the exposure the specimen was kept at − 60°C in order to avoid any further transformation. Immediately after quenching, and with a 1½ hours' exposure, no streaks were visible on the x-ray photograph. Ageing for

a further 2 hours followed by a 2 hour exposure again produced no streaks. A 6 hour exposure after a further ageing of 5 hours produced very weak streaks upon the photograph, and these became much more prominent after another 4 days of ageing. Experiments were also carried out at 100°C, and at this temperature it was found that precipitate was very faintly discernible in the grain boundaries of the material after 120 days' ageing. Streaks were also visible upon the x-ray pictures and were found to be stronger and more well defined than those which had been found in the samples aged at 25°C. At 150°C the first signs of precipitation could be seen inside the grains of material aged for 24 hours, and the amount steadily increased with time of ageing until after 32 days it was present in considerable quantity, although even after that time the particles were extremely small so that their shape was not clearly discernible. The general effect seemed to be that the particles occurred upon parallel lines throughout the crystals, although there were some local aggregations. Streaks were visible upon the x-ray transmission diagrams, but their appearance varied with the time of ageing. After one minute they resembled those seen in the samples aged at 100°C, but with longer times they became narrower, and after still longer, spots corresponding to θ' appeared, but there was no sign of any spots due to $CuAl_2$. At 200°C the precipitate was visible after 2 hours, and after 24 hours appeared as thin particles forming clusters of various sizes. Streaks were visible upon the x-ray photographs, and spots corresponding to the θ' phase first occurred after 2 hours and became more pronounced with longer ageing. No signs of $CuAl_2$ could be seen, even after 41 days. At 250°C precipitate could be seen after five minutes; later well formed needles appeared, followed by small rounded particles and after 52 days the needles had thickened considerably. The orientation of the needles was different in different crystals, appearing to be in two directions in some of the crystals and in three directions in others, that is, in conformity with cubic structure. Streaks did not appear on any of the x-ray photographs, but spots due to θ' appeared faintly after five minutes and became more intense as ageing was continued. Weak lines due to $CuAl_2$ appeared after 32 days, and their appearance was such as to indicate the presence of very small precipitate particles, which were not completely randomly orientated. At 300°C precipitate was seen after thirty seconds and rapidly developed into needles. No

streaks were visible on the x-ray photographs, but weak spots due to θ' appeared after thirty seconds, and $CuAl_2$ was first seen after one hour.

The streaks in the x-ray photographs were shown to be due to small diffracting areas within the metallic crystals, and the size and thickness of the regions could be estimated by analysing the intensities of the streaks. It appeared that these small elements were dispersed at random throughout the lattice and were composed of atoms which differed from those of the main lattice, and assumed by Calvet, Jacquet, and Guinier to be copper atoms. Guinier and Preston had shown previously that the small elements were parallel to the cube planes of the lattice. It was therefore argued that an alloy that showed the streaks possessed a lattice structure similar to that of aluminium but having a slightly different parameter, in which the copper atoms were not situated in the normal random manner characteristic of a solid solution but were concentrated on certain crystallographic planes, namely the cube planes. The dimensions of the randomly orientated groupings were estimated as very small and less than 50Å at 25°C, and 150 to 200Å at 100°C. At 150°C the initial size would be the same as at 100°C but would grow to 600Å after a time. At 200°C the groups would reach a size of 800Å. The larger aggregates also appeared to have an increased thickness, a fact which could be deduced from the different density of the streaks in such cases. As the aggregates became larger there would be a tendency to acquire a more stable configuration, so that they would gradually form a crystal which would be flat and orientated at first with respect to the solid solution. This would correspond to the θ' phase.

In connection with the metallographic examination, Calvet, Jacquet, and Guinier mentioned the striated structures reported by Fink and Smith which had been attributed to localized precipitation upon planes that had suffered plastic deformation. It was found possible to reproduce such striated appearances in specimens by varying the conditions of electrolytic polishing, even using specimens which had been aged only for short periods. It was concluded that the striations did not afford evidence of precipitation but appeared at certain stages in the ageing process according to the polishing treatment. It was possible to obtain from alloys which had partly aged specimens which did not give such striations, although the polishing conditions required careful control. In such cases the x-ray diagrams showed the characteristic

streaks but no evidence of precipitation of θ′ or θ. With alloys aged to a greater extent the striations could be seen whatever the polishing treatment, and examination under high power revealed an actual precipitate. At this stage of ageing the x-ray diagrams also showed the presence of θ′. The striations thus appeared to be an etching effect not necessarily connected with precipitation and for which quenching stresses could well have been responsible.

The hardness of the specimens was also determined as a function of the time of ageing, and maxima were obtained at 100°C and 150°C. At room temperature hardening was appreciable, although there was no evidence of the formation of aggregates. However, the hardening might be due to the beginning of the formation of aggregates. The softening of the alloys at high temperatures appeared to be associated with transformation of the θ′ into θ, a process which appeared to be complete at 300°C. At intermediate temperatures the aggregates and the θ′ existed together. A logarithmic relationship appeared to hold between the time at which precipitate first became visible under the microscope and the reciprocal of the absolute temperature of ageing.

This work was a considerable advance in the theory of ageing, as it showed without much doubt that a definite pre-precipitation segregation of solute atoms occurred in the case of the aluminium-copper alloys, very similar to that suggested in Merica's original knot theory.

THE MECHANISM OF PRECIPITATION

In 1939 MEHL and JETTER[47] analysed in some detail the nature of the precipitation process and the evidence for and against the precipitation and pre-precipitation theories of ageing. The formation of a particle of precipitate was first considered by analogy with the VOLMER and WEBER[34] theory of nucleation, which had been developed for the condensation of a vapour to a liquid and indicated that nucleation could only begin when chance fluctuations in the vapour provided a nucleus greater than a certain critical size. This critical size arose owing to the fact that the formation of a nucleus of liquid required the formation of an interface between the nucleus and the vapour, involving an increase in energy which had to be provided by the vapour to liquid transformation. Only if the free energy of the new phase was less than that of the original phase could growth of the nucleus take place. The

theory indicated that smaller nuclei would be stable the higher the degree of supersaturation and undercooling. At the equilibrium temperature for transformation the stable nucleus would be very large, and with decreasing temperature it would become smaller.

This type of theory was applied by Becker to precipitation from metallic solid solutions, and similar ideas were found to be true. For each degree of supersaturation in the solid solution there was found to be a definite size of nucleus which must be formed by local fluctuations in concentration before growth of the precipitating phase could take place. The rate of formation of nuclei was also calculated for the beginning of precipitation, and it was found that the rate of nucleation was very small both at the equilibrium temperature and also at temperatures considerably below this, but reached a maximum in between.

The calculation of stable nuclei sizes proved more difficult in metal systems than for the vapour to liquid transformation, since the nucleus is crystalline and hence the work required to generate an interface differs for different crystal faces.

Mehl and Jetter pointed out that Fink and Smith[48] had assumed the distribution of atoms in solid solution to be statistically random and used the Poisson equation to calculate the probability of the occurrence of various numbers of solute B atoms in certain size groups in a solid solution composed of A and B atoms. They had applied the calculation to the finding of the number of chance groupings of different sizes formed in a solid solution which would have the correct number of copper and aluminium atoms to form $CuAl_2$. The total number of such groupings was found to be very large, but the number of any one size decreased rapidly as the size of group was increased.

The changes expected to occur during precipitation were considered in detail by Mehl and Jetter. They pointed out that precipitation could be of several types, depending upon whether the matrix concentration of solute atoms was uniformly decreased during ageing or whether the depletion was complete in one portion of the crystal before it had begun elsewhere. These two types were referred to as continuous general and localized respectively, and a third type, namely continuous localized, was described where the solute concentration in the matrix diminished continuously in certain regions but discontinuously as a whole.

The changes of property expected to occur during what was termed ideal precipitation were then discussed. In such precipitation it was

assumed that the changes taking place were those accompanying the transformation from a uniform solid solution to a heterogeneous mixture, and that the initial and final properties could be obtained by considering the properties of solid solutions and mixtures. In such an ideal case no account was to be taken of any effects which might arise in the final product owing to the particles being very small or to the interfaces created between the particles and the matrix *etc.* This and other effects would arise in practice, however, and modify the actual changes observed on precipitation.

The changes in microstructure, electrical conductivity, density, magnetic properties, mechanical properties, and x-ray diffraction effects which would take place during ideal precipitation were compared with those which were observed in practical precipitating systems. In connection with microscopical examination the marked difference in type of precipitation in various systems was mentioned in some detail. Alloys of aluminium and copper showed no preferential grain boundary precipitation although the precipitation might become localized due to strain. In copper-silver alloys rich in copper, however, precipitation appeared to take place preferentially at the grain boundaries and spread inward in an advancing wave, the decomposition reaction taking place only at the interface between the transformed and untransformed regions. Mehl and Jetter regarded continuous precipitation as the simple type and discontinuous as being anomalous, and mentioned that the nickel-beryllium system could show either type, depending upon the solute concentration and temperature.

It was pointed out that microstructural changes had been found to occur at the same time as hardening in several systems such as copper-beryllium, iron-tungsten, and copper-silver, but for aluminium-copper and duralumin no precipitate had been observed at the point of maximum hardness although Fink and Smith claimed that their striated structures were evidence of submicroscopic precipitation. The localized nature of the precipitate indicated, however, that it had been formed preferentially owing to an accelerating factor such as strain.

It was pointed out that the difficulty arising from Fink and Smith's work, which had been discussed previously by other investigators, was that although traces of precipitate had been found from the beginning of ageing, this in itself was not proof that the hardening and

other changes were caused by the precipitation. Pre-precipitation changes taking place in the matrix at the same time as the localized precipitation was occurring may well have been the factors causing hardening. Mehl and Jetter held, however, that the absence of visible precipitate in the undistorted region of the grains could not be taken as conclusive evidence that precipitate was not present, as it might be of a submicroscopic character. The action of the distorted regions would be to concentrate the precipitate and make it visible.

In any discussion of theories of ageing it is necessary to arrive at some definition of a nucleus of precipitate if any real distinction is to be drawn between pre-precipitation and precipitation ideas. Mehl and Jetter considered that a nucleus of precipitate had formed when the atoms of a solid solution had begun the movements which would eventually lead to the production of the equilibrium phase demanded by the phase diagram. The Guinier-Preston zones which were known to precede the formation of θ' and $CuAl_2$ in aluminium-copper alloys were regarded in this light as being true nuclei in the precipitation process. If this was not so, then it was difficult to see why the zones should appear in a definite plate-like form along those very planes upon which the θ' and finally $CuAl_2$ would form. If they were not true nuclei, then it would be expected that they would take the form of random shapes, similar to those caused by concentration fluctuations. Furthermore, the variation of the size of Guinier-Preston zones with temperature as determined experimentally was the same as indicated by the theoretical considerations which had been developed for nucleation, and the actual sizes were within the expected range. It was therefore considered that Guinier-Preston zones were nuclei of precipitate, and that once the formation of such zones had begun it might also be said that precipitation had begun.

Working upon such an assumption, Mehl and Jetter explained the changes in the various properties which took place during ageing. The observations of Fink and Smith on the striated structures produced in aluminium-copper alloys were assumed to result from the formation of Guinier-Preston zones, which would effect the etching characteristics of the alloy. These regions would then subsequently grow and form θ' and θ so that definite particles would become visible under the microscope. Resistance changes could be explained by Mott's suggestion regarding the diffracting action of small particles on the

electrons. It was considered, however, that other factors would be of importance in determining the resistance, such as the state of strain present in the alloy, particularly at the junction between the precipitate particle and the matrix. This strain would be the greatest when the precipitate particles were smallest, since the total amount of strain present would be dependent upon the relative surface areas and volumes of the particles. Small particles formed at low temperatures would thus give the largest strain, an additional effect contributing to the initial increase of resistance during room temperature ageing.

The inability of x-rays to detect parameter changes was regarded as being in no way incompatible with the idea that Guinier-Preston zones were actual precipitate particles. Parameter measurements had been shown to be no criterion of precipitation, and certain small anomalies in parameter measurements could probably be put down to strain.

The volume changes during ageing of aluminium-copper alloys at room temperature or at high temperatures could also be explained. Fink and Smith had pointed out that by the formation of θ' of greater density without a change in volume of the matrix a contraction could result. At higher temperatures the nuclei would become larger and less numerous, and their growth by diffusion from the matrix, with a corresponding increase in volume of the matrix, would lead to an overall expansion. The interfacial regions were also regarded as regions of low density which could perhaps also contribute to the observed volume changes.

An interesting anomaly to which reference was made was that of the copper-iron system, where despite the fact that iron was known to precipitate during ageing the alloys did not become ferromagnetic. It was thought that this could be due to two effects: either the ferromagnetism would be difficult to detect when the precipitate particles were fine, or alternatively the lattice structure of the iron was altered during precipitation by conforming to a certain extent to the matrix lattice. This straining of the iron lattice would alter its magnetic properties, and it would not be until the lattice reverted to its true form by increased growth or mechanical shock that the expected magnetic properties would be observed.

Mechanical properties were not regarded by Mehl and Jetter as being of much use when attempting to formulate theories of ageing, owing to their complex nature and the difficulty of interpretation.

They pointed out that double ageing peaks were difficult to reproduce and that the explanation of Fink and Smith seemed to cover all the experimental data. Nevertheless there was a possibility that the succession of stages involved in precipitation, that is, Guinier-Preston zones, θ′, and finally $CuAl_2$, would be in themselves sufficient to give more than one maximum on a property/time curve. All these stages were now regarded as post-precipitation rather than pre-precipitation stages so that their presence or otherwise did not invalidate the theory that ageing was due to precipitation.

Discussion of Mehl and Jetter's views indicated that the gap was closing between those who believed the pre-precipitation processes were responsible for ageing changes and those who believed actual precipitation to be the most important. MERICA[49] pointed out that the original supporters of the precipitation theory were those who argued that there was no evidence during the room temperature ageing of duralumin of anything present in the structure which answered to the known tests for $CuAl_2$. The changes causing ageing must have been of a homogeneous nature which did not involve any local lattice alteration of a definitely discontinuous kind. It appeared from Mehl and Jetter's paper, however, that precipitation was assumed to take place in stages, the first stage being the formation of Guinier-Preston zones. The structure of the zones was not known, but could be either very similar to the aluminium lattice or to the θ′. Since the zones were intimately associated with the hardening, then it seemed that the correct theory of hardening was half-way between the pre-precipitation and precipitation theories, as the Guinier-Preston zones did not appear to be homogeneous with the aluminium lattice in the strictest sense, nor did they appear definitely to be particles of tetragonal $CuAl_2$.

COHEN[50] considered that much of the controversy in ageing theory was only a question of terminology. Originally the idea of a precipitate particle was thought to involve a definite phase boundary with the matrix lattice, even though it was realized that prior changes would go on in the solid solution. Mehl and Jetter, however, were using precipitation in a very much broader sense to cover both initial and final stages. It seemed, therefore, that there was very little significant difference between Merica's original ideas of knots and the present idea of a Guinier-Preston zone nucleus. In both cases the boundary with the matrix would be equally indistinct. The important

consideration, according to Cohen, was not whether one spoke in terms of knots or precipitated nuclei, but with what changes one associated the alteration in properties on ageing. Practically all were in agreement that it was the lattice distortion caused by the straining of atomic bonds which resulted in the hardness, density, resistance changes *etc.*

Cohen pointed out, however, that despite the partial reconciliation it was untrue to say that double ageing peaks were not of fundamental significance as they could be reproduced quite readily and were not to be attributed to experimental error. There were many examples of them in the literature, and it was important to investigate them in detail as there might well be two stages of hardening during ageing, the first being due to lattice strain and the second to a mechanical keying of the matrix slip planes.

Fink and Smith[51] referred to the structure of Guinier-Preston zones and thought that the zones were platelets of θ′ produced by the formation of a small θ′ nucleus and the subsequent growth of the nucleus along the cube planes of the matrix lattice. The flat method of growth was due to the fact that the atomic planes in contact between the θ′ and the lattice were similar so that the register between the lattices was good, and therefore extension sideways would be the easiest method of growth.

Phillips[52] raised the important point that in considering ageing changes attention should be concentrated not only on the direct effect of the products formed by pre-precipitation or precipitation mechanisms but also to any changes which might be produced in the matrix lattice by reason of the formation of pre-precipitate or precipitate regions.

Mehl and Jetter[53] further discussed their proposed ageing theory and pointed out how little it differed from Merica's original theory; that the distinctions drawn between various theories appeared to hinge upon the question of definitions. This was illustrated by means of the aluminium-copper system where the sequences of changes leading to precipitation were known to be

a the solid solution state,
b Guinier-Preston zone formation,
c θ′ formation, and
d final formation of $CuAl_2$.

Earlier investigators had drawn the demarcation line, so far as the onset of precipitation was concerned, at the beginning of stage *d*,

though since θ′ could be recognized under the microscope as precipitate, probably it would be more accurate to say that the beginning of stage *c* had been used. The main change in opinion which seemed to have resulted was that with the new viewpoint the formation of Guinier-Preston zones was regarded as definite precipitate formation. The anomalies found during ageing were all associated with stages *b* and *c*, which thus resulted in the pre-precipitation *versus* precipitation dispute. The stages during precipitation would, however, be of a continuous nature and nothing very much could be gained by attempting to define rigorously terms such as 'phase' or 'interface'.

One difference which did exist, however, between the ideas of Mehl and Jetter and those who favoured pre-precipitation ideas such as knot formation lay in the fact that the latter workers postulated that diffusion of solute atoms to certain crystal planes took place to form the knots or Guinier-Preston zones, whereas the former did not think any special diffusion processes were necessary but that the localized aggregations would always exist in solid solutions as a result of normal fluctuations in concentration.

Mehl and Jetter did a very great deal to clarify the theoretical basis of ageing changes. The outcome of their paper was in effect that the dispute between the precipitation and pre-precipitation advocates was largely resolved as it became clear that it had become merely a matter of terminology. The changes leading from a random supersaturated solid solution to a final stage, consisting of a mixture of precipitate particles and solid solution, demanded by the equilibrium diagram, had been described in detail for the aluminium-copper alloys. It was realized that the important task from the point of view of ageing theory was to clarify which particular stages in the sequence of steps leading to precipitation were responsible for the observed experimental changes in properties. It appeared that the formation of Guinier-Preston zones and θ′ were the steps which led to the most marked changes in properties, but this was by no means certain and could merit further investigation. Furthermore, though the changes taking place prior to precipitation were well known in aluminium-copper alloys, it was not known whether similar changes always preceded precipitation in other systems. Fink and Smith[54] had shown that for aluminium-magnesium alloys the phase which first precipitated when the alloys were aged at low temperatures was not the equilibrium

phase but one which had a slightly different crystal structure. They had pointed out that there was a possibility that for all precipitate formation the initial precipitate that formed was not the equilibrium phase, as predicted from the equilibrium diagram, but a transition non-equilibrium phase which later transformed to the equilibrium

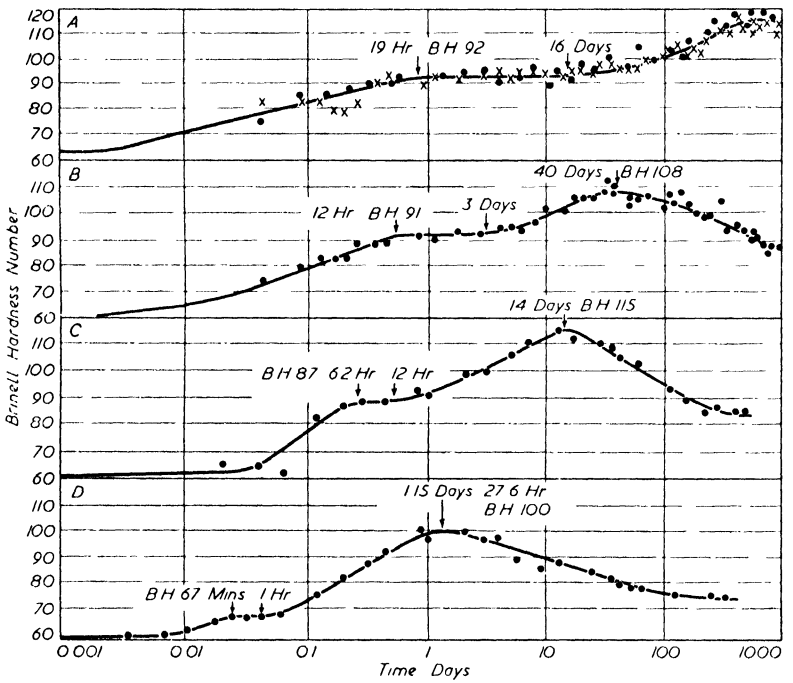


Figure 11. Hardness/time curves of an aluminium-copper alloy (Gayler)

structure. No evidence was brought forward, however, which indicated the formation of regions in the lattice similar to the Guinier-Preston zones in the aluminium-copper system.

MICROSTRUCTURAL INVESTIGATIONS

The microstructural evidence obtained during ageing, particularly of aluminium-copper alloys, was also rather unsatisfactory because the only work which had revealed early traces of precipitate was that of Fink and Smith, and this work could be criticized because of the strain which was present in their specimens. In 1940, however, GAYLER,[56] and GAYLER and PARKHOUSE[55] carried out a detailed microscopical examination of the ageing of a 4 per cent aluminium-

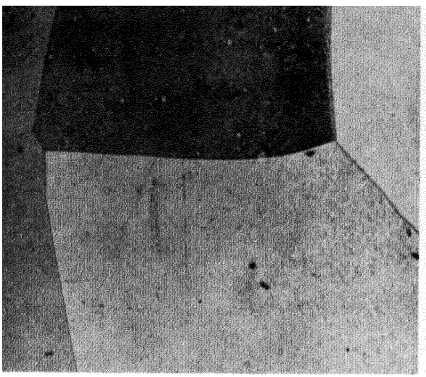


Plate XVII. Aged 600 days at room temperature. Etched 1 minute with 25 per cent HNO_3 *at 70°C and 1 minute with Keller's reagent* (× *150*)

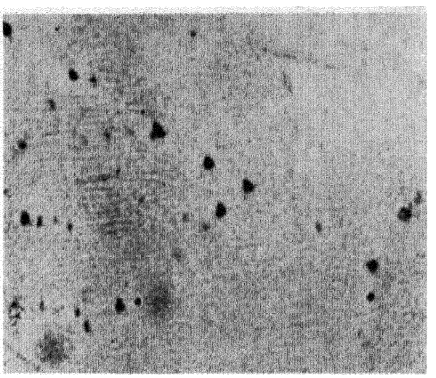


Plate XVIII. As Plate XVII (× *2500*)

Plates XVII and XVIII. Showing staining of grains and etch pits

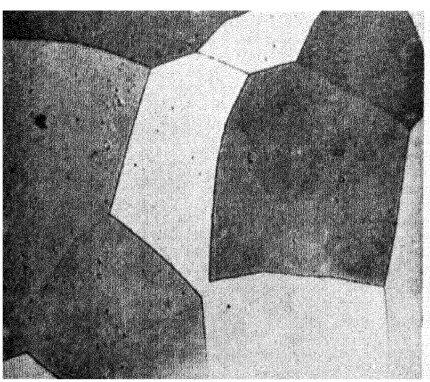


Plate XIX. Aged 230 days at 50°C. Etched as Plate XVII (× *150*)

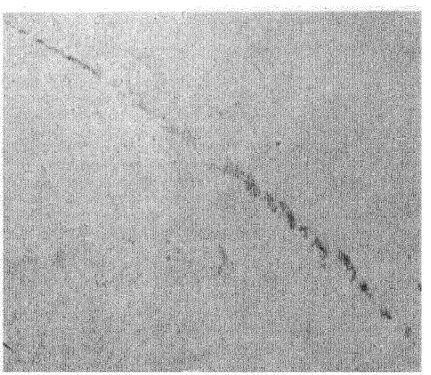


Plate XX. As Plate XIX. Only etched with 25 per cent HNO_3 *at 70°C* (× *1500*)

Plates XIX and XX. Showing staining of grains and etch pits

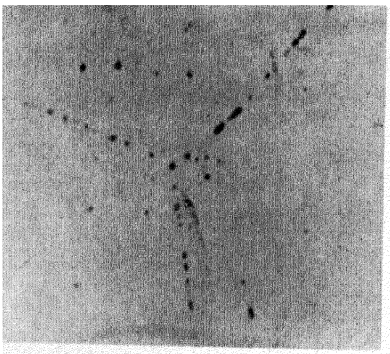


Plate XXI. Aged 2 days at 130°C. Etched 1 minute with 25 per cent HNO_3 *at 70°C* (× 2500)

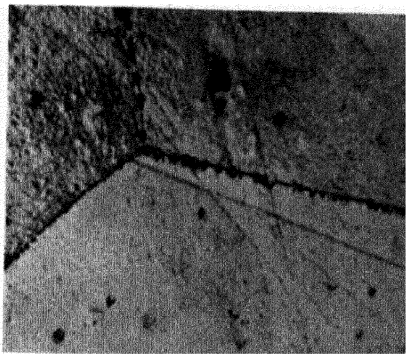


Plate XXII. As Plate XXI but etched as well with Keller's reagent for 1 minute (× 2000)

Plates XXI and XXII. Showing etch pits but no precipitate

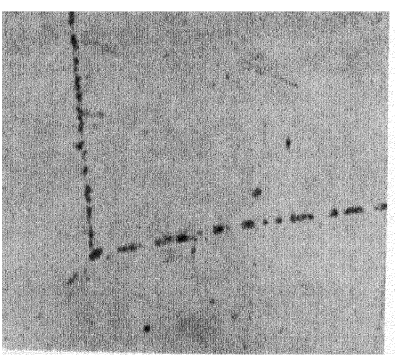


Plate XXIII. Aged 7 days at 130 C. Etched 1 minute with 25 per cent HNO_3 *at 70°C. Showing etch pits* (× 2500)

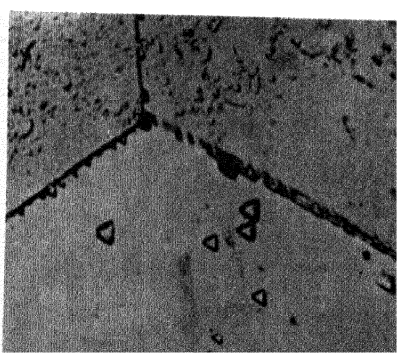


Plate XXIV. As Plate XXIII, but etched as well for 3 minutes with Keller's reagent. Showing etch pits and precipitate (× 2000)

copper alloy; this was supplemented with hardness measurements and a certain amount of x-ray work.

Ageing at room temperature was continued for 600 days, and no increase in hardness was observed until after eight days. It appeared, however, from this and experiments at higher temperatures that the ageing took place in the stages that had been postulated earlier (*Figure 11*). At ageing temperatures up to 50°C only the first stage was found to be present, whereas between 100°C and 190°C two maxima were found. At still higher temperatures it was not possible to observe the first hardness maximum and only the second was found. Ageing at the quenching temperature resulted in an immediate softening to the as quenched state.

On the time/hardness curves flat portions were found between the first and second hardness maxima, which extended for a period of time that became greater at lower temperatures of ageing. If the times between the first hardness maximum and the beginning of the increase due to the second hardening were plotted against the reciprocal of the absolute temperature, a straight line was given. Similarly, when the time taken to soften by 10 Brinell units immediately after the second hardness maximum had been reached was plotted against the reciprocal of the absolute temperature, another straight line resulted. From such curves, results of considerable practical value, relating to the time taken for various stages of the ageing to be reached or for various amounts of softening to set in, could be obtained. Plots of the logarithm of the time to attain the second maximum hardness and the reciprocal of the absolute temperature were also made, and a linear relationship was found for temperatures between 100°C and 400°C. This temperature range was limited, since on ageing at the temperature of quenching no increase in hardness took place, and on ageing at low temperatures the second maximum would not be observed because the ageing would not proceed so far. No such linearity existed at all when the time to reach the first maximum hardness was plotted; the curve showed a marked inflection.

A plot of the maximum for each stage against the ageing temperature showed that, within the limits of the experiment, the maximum hardness during the first stage of ageing increased up to an ageing temperature of about 50°C and then decreased. For the second stage the maximum hardness increased with the ageing temperature until a

maximum was reached at about 100°C to 150°C. It was thought that the curves should become asymptotic to the temperature axis, as no ageing would occur at temperatures around 0°C. The probable form of the curves would be, in each case, an increase with temperature up to a maximum followed by a decrease.

In the microscopic work carried out by Gayler special attention was paid to the ageing temperatures below 150°C as it was believed that below that temperature the first stage of ageing played a predominant part. The results of previous workers on ageing at low temperatures were conflicting. Fink and Smith claimed to have shown definitely the presence of precipitate, whereas Calvet, Jacquet, and Guinier found no precipitate on ageing at room temperature for as long as 340 days. It appeared, however, that Fink and Smith worked with specimens that had received a considerable amount of plastic deformation during the quenching, and that the etching effects which they found were connected up with this rather than the presence of actual precipitate particles. The striations probably appeared because the surfaces of the crystals were deformed, and the experiments of Calvet, Jacquet, and Guinier on the production of such striations tended to confirm the point. High magnifications were used by Gayler, generally $\times$ 2,500, and the etching was carried out with 25 per cent nitric acid alone at 70°C or followed by Keller's etch.

At room temperature no precipitate was visible even after 600 days' ageing. On etching the grains of the solid solution were not all of the same colour, some appearing to be much more heavily stained than others (*Plate XVII*). This variation in the grain staining appeared to be associated with the first stage of the ageing, and it could be observed in specimens which had been aged for only a short time at room temperature *e.g.* 15 minutes. On further ageing, however, other changes became visible. Nitric acid, followed by Keller's etch, caused etch pits to appear in the grain boundaries and also within the grains (*Plate XVIII*). This was only to be seen in specimens which had been aged for a considerable time at room temperature. No striae such as described by Fink and Smith were seen. At higher ageing temperatures again no precipitate was seen, but the preferential staining of certain grains was still noticed (*Plate XIX*), and the grain boundaries again appeared to have become more reactive as etch pits were produced by nitric acid after ageing for 250 days (*Plate XX*). The pits

appeared as either rhombohedral shapes or small round areas. At 130°C ageing produced etch pits after only 2 days (*Plates XXI, XXII*) which were of a definite triangular shape and were present in the interior of the grains as well as the boundaries. No precipitate was observed in any of the specimens aged at the above temperatures. It thus appeared that the formation of the etch pits was connected with the initial hardening of the alloy and with the portion of the hardness/time curve connecting the initial hardness maximum with the second maximum. The specimen which had been aged at 130°C for two days was aged for a further five days, when the increase in hardness due to the second stage of the ageing should have begun to appear. No precipitate was visible in the unetched condition, but when etched with nitric acid etch pits were seen (*Plate XXIII*) and etching with nitric acid and Keller's etch showed marked precipitation (*Plates XXIV, XXV*). The preferential staining of the grains was still present; precipitation was not visible in the unstained grains though in some cases etching pits were found (*Plate XXVI*). No precipitate seemed to be present in the grain boundaries, although there was a considerable number of etching pits. Further ageing at 130°C resulted in a still greater amount of precipitate becoming visible, even in the unstained grains. After fifty days ageing a further effect was distinguished: a small amount of a white material appeared in the grain boundaries, associated with etching pits (*Plates XXVII, XXVIII*). The alloy had been aged to a point just past the maximum hardness. It was thought that the white material could possibly be the intermediate phase of Wassermann and Weerts. In any case it was not $CuAl_2$, as its etching characteristics showed. The x-ray analysis of specimens aged for 50 days at 130°C showed no shift of the lines. This indicated that precipitation proper did not take place until after the second hardness maximum. At higher temperatures of ageing, above 190°C, the changes due to the first stage were masked by general effects. It had been shown previously by Wassermann and Calvet, Jacquet, and Guinier that at such high temperatures the formation of θ' preceded the formation of θ. It was possible, according to Gayler, to detect the change of θ' into θ by means of microscopical examination after suitable etching, and it was noticed that upon etching a definite boundary always appeared between the θ and the aluminium matrix, while in the case of θ' there was no such definite boundary, the phase seeming to run on into the aluminium

lattice (*Plate XXIX*). This was regarded as evidence in support of the conclusion reached by PRESTON[57] that θ′ was related to the aluminium lattice, so that the (100) planes of the aluminium lattice were in contact with (001) planes of the θ′. A plane of polish which deviated only a small amount from being parallel to the (100) plane of the aluminium would show plates of the intermediate phase, some edges of which would stand out in relief from the matrix due to hardness differences, while other portions would run on into the aluminium lattice due to the small difference in orientation.

Gayler showed that the staining of the grains in the initial stages of ageing was related to the orientation of the grains with respect to the plane of polish of the specimen. The deepest staining occurred when the (100) planes were parallel to the plane of polish, while unstained grains appeared to have the (110) planes parallel to the plane of polish. Intermediate degrees of staining were probably due to orientations lying between these two. The staining was due in all probability to the fact that copper atoms segregated upon the (100) planes of the alloy and thus affected the action of the etching reagent.

This metallographic work indicated that the various stages of ageing in aluminium-copper alloys were characterized by different responses to etching reagents. During the first stage, up to the first maximum hardness, no visible precipitation was seen, but the preferential staining was regarded as characteristic. On the flat portion between the first and second maximum the chemical reactivity of the interior of the grains and the grain boundaries increased, as evidenced by the formation of etch pits. The increase to the second hardness maximum was associated with the appearance of θ′ within the grains, but no precipitation appeared in the grain boundaries until after the second maximum hardness was attained when the white material appeared. After the second maximum hardness the change from θ′ to $CuAl_2$ took place gradually. It was questionable whether the phase appearing in the grains before the second maximum hardness was reached was to be regarded as definite precipitate or not. The fact that it was still attached to the aluminium lattice was regarded as evidence against this.

An extension of the work, with further detailed microscopical examinations,[58] was reported in 1946. The changes in microstructure which occurred on ageing at 130°C up to the second increase in hardness had already been dealt with, but the white material which had been revealed

had not been studied in detail. To investigate whether its growth took place on prolonged ageing, specimens were aged for 50, 80, 200, and 400 days at 130°C. Etching with hot 25 per cent nitric acid followed by ½ per cent hydrofluoric acid revealed the material (*Plates XXX, XXXI*), but no appreciable growth in the boundaries appeared to have taken place in the 200 day specimen as compared with the 50 day.

An increased amount of the material was, however, present in the interior of the grains. Etching pits were found to be associated with it, and it was thought that these indicated segregation of copper to form copper rich areas. In support of this it was pointed out that 25 per cent nitric acid attacked copper rich areas but not grain boundaries in pure aluminium or a dilute solid solution of copper in aluminium. Half per cent hydrofluoric acid, on the other hand, attacked grain boundaries in pure aluminium or a dilute solid solution of copper in aluminium, but attacked $CuAl_2$ only very slightly. The etching pits were therefore taken to be the traces of the copper rich regions. Simultaneously with the formation of such regions, adjacent areas nearly copper free would be formed.

In the specimen which had been aged for 400 days and which was past the second hardness maximum it was found that on etching with half per cent hydrofluoric acid alone the white material was revealed, and also small areas in the boundaries and grain interiors were darkened. In previous specimens the hydrofluoric acid had revealed only the former phenomenon, and it was thought that the new areas which were being attacked were particles of θ'. Previous to 1940 the light material was thought to be θ', but it was now postulated that it was the aluminium-copper solid solution stable at the temperature of ageing. The material was given the name crystallite. It appeared that as the time of ageing was prolonged there would be an increase in the number of copper rich areas, and it was postulated that when these reached a critical size the stresses generated in the surrounding matrix would be relieved by local recrystallization of the matrix with consequent formation of crystallites. The number of crystallites would therefore increase with time of ageing. They did not grow but merely increased in number with increased time and temperature of ageing.

The rate of cooling from the solution treatment temperature was found to exert an influence on the microstructure obtained on subsequent ageing. Decreasing the rate of cooling led to the formation of

an increasing amount of crystallites during ageing. This was assumed to be due to slower rates of cooling, allowing segregation of copper atoms to occur. This would give an explanation of Preston's observation that streaks were to be seen on x-ray photographs taken immediately after quenching, for these streaks would be due to the slight segregation of copper atoms during quenching.

The assumption that nitric acid attacked copper rich areas was confirmed by using a selective tarnish of hydrogen sulphide. An air cooled specimen was polished and exposed to very dilute dry hydrogen sulphide after ageing for 426 days at room temperature. No change was observed after 41 days, but after 3½ months the grain boundaries were seen to have been etched up and the solid solution grains were selectively tarnished (*Plate XXXII*); probably this arose from their differing orientations, grains with (100) planes parallel to the surface being most heavily stained. Crystallites were seen in the grain boundaries and the adjacent darkened areas were taken to be copper rich areas which had been selectively tarnished. These darkened areas occurred in those positions in which etch pits appeared when the specimens were etched with nitric acid. It appeared, however, that the nitric acid etch could reveal the copper rich regions when they were submicroscopic, as one specimen showed no darkened regions when exposed to hydrogen sulphide but showed pits with nitric acid. It was ascertained that $CuAl_2$ was unaffected by hydrogen sulphide, and also that θ' and $CuAl_2$ in the transition stage were not darkened.

Slower rates of cooling prior to ageing were found to result in longer time delays before an increase in hardness was observed. This was thought to be due to localized segregation taking place to a greater extent during slower cooling.

The segregates formed on slow cooling would be large, and as soon as ageing began they would probably precipitate out, thus producing a decrease in hardness which would balance the tendency to hardening arising from the formation of new copper rich areas. The initial incubation period therefore represented the time during which the aggregates formed during cooling were growing past their critical size and then precipitating.

The time/hardness curves obtained on ageing specimens at room temperature after different rates of cooling were all found to be different, due to the differences in the amounts of segregation that had taken place.

On ageing at 130°C, however, the curves were all found to be similar after the first twelve hours or so of ageing. This was explained by assuming the initial differences to be due, as at room temperature, to the size of the aggregates formed on cooling. The secondary portions of the curves would become similar, due to a resolution of the copper rich areas followed by rediffusion to form new aggregates. The resolution was confirmed by ageing a slowly cooled specimen for five hours at 130°C and then comparing the new microstructure with the original, after exposing both to hydrogen sulphide. No stained patches were seen in the specimen that had been aged at 130°C though they were present in the original specimen, thus indicating that resolution of the copper rich areas had taken place. Gayler pointed out that the presence of double ageing peaks on the curves at 130°C after the various rates of cooling was a contradiction of Fink and Smith's statement that the first ageing peak was due to plastic deformation and that it could be eliminated by cooling at a slow rate.

It was concluded that the various stages in the ageing of aluminium-copper alloys could be related to changes in behaviour of the material on etching. The initial increase in hardness was not accompanied by an etching change, but the flat portion between the first and second maximum marked the point at which minute etching pits were formed when the specimens were etched in nitric acid. These were not shown by hydrogen sulphide and were assumed to be submicroscopic copper rich regions. During the increase to the second hardness maximum the copper rich regions could be detected by hydrogen sulphide, which indicated that they had grown, and furthermore etching with hydrogen fluoride revealed the crystallites of aluminium solid solution. As the hardness increased to a maximum the number of crystallites increased, and after the maximum a new phase associated with the crystallites could be detected by etching with hydrofluoric acid. The needle-like particles of the new phase were θ', which subsequently transformed to $CuAl_2$.

GAYLER'S THEORY

Gayler enumerated the changes in properties associated with the various stages of ageing and advanced a complete theory in explanation. In the initial stages, segregation of copper atoms was thought to take place on the (100) planes of the solid solution lattice. The segregation

would cause an increase in density, due to the local contraction of the lattice, the copper atoms being smaller than the aluminium atoms. A change in lattice parameter would also be expected if the views of Guinier and Preston were accepted, but no change had been found, due perhaps to the localized nature of the segregation. Strain hardening would be produced by the segregation and cause an increase in the electrical resistance and the hardness. It was postulated that the strain hardening hindered the process of diffusion, so that the migration of copper atoms would cease. A certain amount of strain relief would come about, however, by the precipitation of submicroscopic particles of copper rich aggregates, and of submicroscopic crystallites of the solid solution stable at the ageing temperature. The local copper segregates would be increased in size by localized diffusion until they reached a critical size, at which point they would break away from the matrix lattice, relieving strain and enabling further diffusion to take place. The flat stage on the ageing curve would then appear, due to the balancing between diffusion of atoms leading to segregation and strain hardening, and precipitation which would cause a relief of strain. At this stage the electrical resistance would decrease. The process of formation and precipitation of segregates would go on and at the same time the first precipitated copper rich particles would grow, leading to a second increase in hardness, thus producing the second maximum.

The next stage would be the reorientation of copper atoms to form θ', and it was suggested that this would only take place in strain free regions, since no θ' could be detected microscopically, even under high magnifications, until some time after the second hardness maximum, nor had θ' been observed on x-ray photographs until the ageing was well advanced. Furthermore θ' was known to separate directly from the lattice at high ageing temperatures, where the lattice would be practically strain free, and θ' particles were first observed associated with crystallites which would be strain free regions.

The change from θ' to $CuAl_2$ was thought to take place slowly, depending upon the temperature and time of ageing, but it would not lead to increase in hardness.

Gayler therefore concluded that ageing was due to the formation and precipitation of copper or copper rich aggregates, accompanied by the formation of crystals of the solid solution stable at the ageing

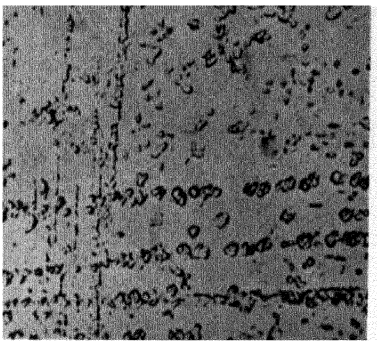


Plate XXV. As Plate XXIV (× 2000)

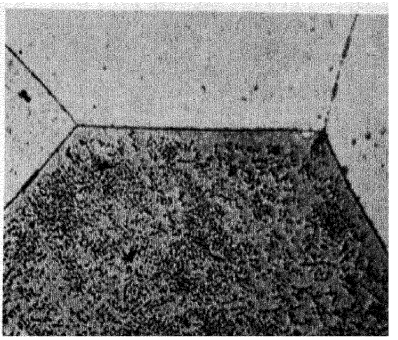


Plate XXVI. Aged 14 days at 130°C. Etched 1 minute with 25 per cent HNO_3 *at 70°C and 3 minutes with Keller's reagent* (× 350)

Plates XXV and XXVI. Showing precipitation

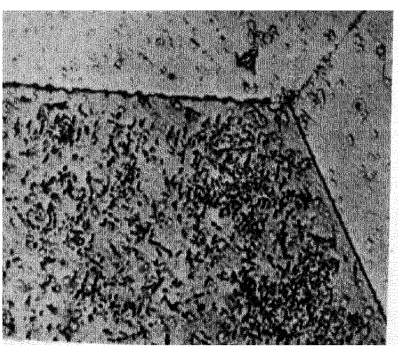


Plate XXVII. Aged 50 days at 130°C. Etched 1 minute with 25 per cent HNO_3 *at 70°C and 1 minute with Keller's reagent* (× 750)

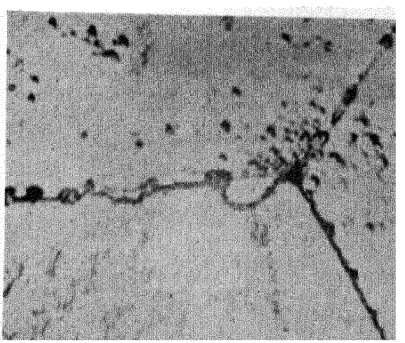


Plate XXVIII. As Plate XXVII (× 2750)

Plates XXVII and XXVIII. Showing traces of the white material

Plates XVII to XXVIII. Photomicrographs of a 4 per cent aluminium-copper alloy (Gayler)

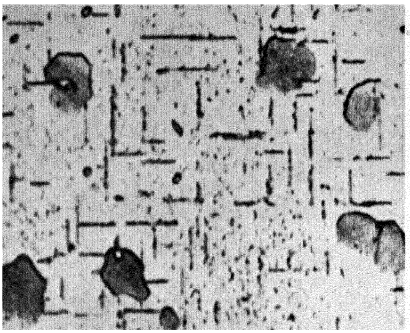


Plate XXIX. 4 per cent copper, 0·3 per cent iron alloy aged 1½ hours at 350°C. Swabbed with ½ per cent HF. *Showing plates of* θ′ (× 2500)

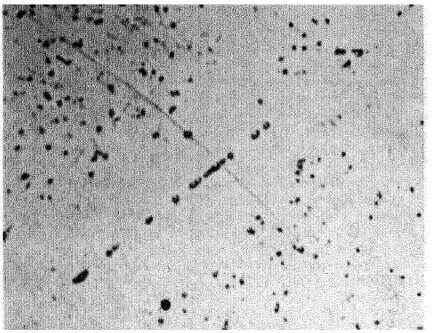


Plate XXX. Aged 80 days at 130 C. Etched 1 minute in 25 per cent HNO_3 *at 70°C. Showing etching pits* (× 2000)

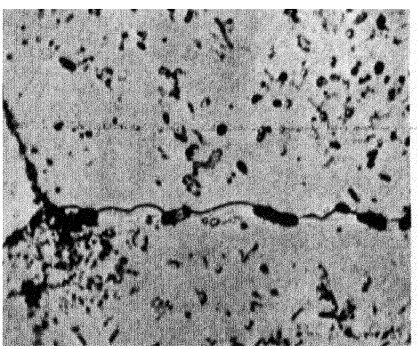


Plate XXXI. Aged 80 days at 130 C. Etched 1 minute in 25 per cent HNO_3 *at 70°C, and then swabbed with ½ per cent* HF. *Showing the white material in grains and boundaries and associated with etching pits* (× 2000)

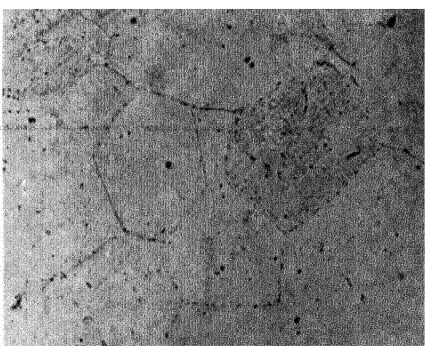


Plate XXXII. Air cooled specimen aged 543 days at room temperature. Exposed to trace of dry H_2S *for 3½ months. Showing staining of copper rich areas and grain staining* (× 150)

Plates XXX to XXXII. Photomicrographs of a 4 per cent aluminium-copper alloy (Gayler)

temperature. The formation of θ′ did not lead to an increase in hardness, as formerly assumed.

An investigation[59] was also carried out on the effect of working a 4 per cent copper in aluminium alloy, at various stages of the ageing process. Cold working was found to accelerate the rate of ageing to an extent which depended on the amount of cold work given. Thus a specimen which had been quenched and given a 50 per cent reduction before ageing at room temperature was found to increase considerably in hardness immediately after the working, and then to remain constant for a period of at least several days. Microscopic examination, after etching with 25 per cent nitric acid, revealed that the specimen after two months was in an advanced state of ageing corresponding to the end of the 'flat' on the time/hardness curve. Quenching and then cold working prior to ageing at different temperature revealed similar effects. A specimen aged at higher temperatures showed an increase in hardness which corresponded to the second increase on the normal time/hardness curve, the cold working having been sufficient to take the alloy initially to an advanced state of the first stage of the ageing.

It was further demonstrated that when the condition of the alloy before working and the ageing temperature were favourable, the lattice strain resulting from the cold work was relieved by the formation of submicroscopic or microscopic copper rich precipitate particles together with the aluminium solid solution crystallites.

It was concluded, therefore, that the strain relief thought to take place during ageing, namely by the formation of crystallites, might also apply to the relief of strain caused by cold working. Crystallites would be formed in the localized highly strained regions at temperatures which could be below the recrystallization temperature. This idea was substantiated by experimental work which had been carried out by other investigators.

In discussing Gayler's papers, Guinier referred to further work which he had carried out on the initial stages of the aluminium-copper ageing. An intermediate stage had been found between the initial arrangement of copper atoms on the (100) planes and the formation of θ′. The expansion of the segregate of copper atoms was thought to be associated with an internal ordering process, resulting in a kind of superlattice being formed by planes of copper atoms or planes rich in copper atoms alternating with planes of aluminium atoms. The superlattice was

thought to be an individual crystal but to remain imperfect, as evidenced by the diffuse nature of the diffraction spots. Where θ′ appeared its diffraction spots were at first weak but sharp, and superimposed upon the diffraction pattern of the superlattice phase. It was thought that the change to θ′ took place suddenly, and only when the superlattice had reached a certain size and the proportion of copper to aluminium atoms present was roughly one to two. Additional spots were found on the x-ray diagrams which seemed to indicate that the θ′ was in a strained region, a state of affairs which was not in agreement with Gayler's ideas. The latter thought, however, that the presence of small strain free areas might not be detected with x-rays.

ROHNER'S THEORY: LATTICE DEFECTS

Gayler's theory of ageing, like those advanced by practically all other investigators, was based upon the idea that the strain introduced into the lattice by the segregation of solute atoms or the formation of precipitate resulted in the observed changes in properties. In 1947, however, ROHNER[60] put forward a new theory of ageing in aluminium-copper alloys which attributed the changes in properties, in the main, to alterations in the lattice of the solid solution, independently of the effect of segregated regions of copper atoms or precipitate particles. Rohner pointed out that the precipitation theory was unable to explain retrogression effects or the electrical resistance changes. (In fact, theories had been put forward to explain these effects.) According to Rohner, most investigators took the available evidence to indicate that there were two stages in ageing, which followed one another closely or overlapped, and in each of which the mechanisms were different. The second stage was regarded as the precipitation of particles of $CuAl_2$, but the evidence for the nature of the first stage was scanty. Theories for this stage, however, included the rejection of solute atoms from solid solution with the possible formation of molecules and the formation of localized segregate regions of solute atoms along certain lattice planes. Rohner considered that the evidence for the two stages was not very strong, and suggested that there was evidence for two counteracting processes, governing one simple kind of ageing. He differed from most investigators in that he thought it highly improbable that the precipitating phase should itself cause the hardening, considering it more likely that the state of the parent lattice was altered

during ageing, and that this led to the observed changes. In other words, the process leading to the precipitation of $CuAl_2$ would cause an alteration of the solid solution lattice.

Experimentally Rohner followed ageing by means of electrical resistance and yield strength changes, Brinell hardness being rejected as too complex a property to study. In making up the alloys very pure material was used, and silicon was excluded in order to avoid any complications. He used alloys of four per cent copper in aluminium, some of which contained small amounts of manganese and magnesium.

In developing his theory Rohner recognized as proven facts, first that the ageing reaction involved an alteration of the parent lattice, and secondly that ageing must consist of two counteracting processes. It was assumed that solute atoms would migrate from a lattice in which they did not fit very well to the interstitial lattice spaces. Precipitation was regarded as being dependent upon a limited solubility of the solute atoms, *i.e.* a low affinity between the solute atoms and the parent lattice. It was reasoned that any diffusion preceding a rejection or precipitation of solute atoms would take place chiefly by migration into the interstitial spaces of the lattice, hence the solute atom would necessarily have an atomic radius less than that of the solvent atom. Rejection of atoms was assumed to take place at grain boundaries or mosaic element boundaries, where the detachment of solute atoms would involve a smaller change of energy and the solute atoms were postulated to reach the localized areas by diffusion through the interstitial lattice spaces. The migrating solute atom would leave a lattice site vacant. The vacancy thus formed could move through the lattice by virtue of the self diffusion of aluminium atoms, and if it reached a position near a grain boundary it might be filled up and eliminated by the entry of a solvent atom from the boundary. The life of a vacant site would thus be determined by the rate of self diffusion of the solvent atoms. It was thought that vacant sites could also be eliminated by migrating solute atoms moving in from interstitial space. Working upon these assumptions, Rohner derived an expression for the manner in which the concentration of vacant lattice sites changed during ageing at room temperatures. Further assumptions were then made as to the dependency of tensile and resistance values upon the concentration of vacant sites, and equations were developed which would enable the correctness of the reasoning and the assumptions to be checked.

Agreement between theory and experiment was good if it was assumed that the resistance increased directly with the concentration of vacant sites, while the elastic limit increased with the cube root of the concentration. It was pointed out that a vacant site would have a similar effect but more pronounced on the resistance as a stranger atom in solid solution. The rate of increase of resistance was shown to be proportional to the rate of passage of copper atoms from the lattice to the interstitial spaces, that is, nearly proportional to the rate of formation of vacant sites. The theory also accounted for the small rate of decrease in resistance at room temperature after long times of ageing.

To explain the increase of strength during ageing Rohner developed a new theory of hardening, based upon the effect of vacant sites on slip, which led to a formula for the elastic limit very similar to that developed by Bragg for the work hardening of pure metals. According to Rohner this was not surprising, since in ageing the lattice would be broken up into slip elements by the vacant sites in a similar manner to which a lattice was broken up into mosaic elements by cold work. Work hardening and age hardening were thus ascribed to the same fundamental mechanism.

Rohner extended the above conceptions to the more complicated problems of direct ageing at higher temperatures, including retrogression effects which appear where ageing continues at a high temperature after being initiated at room temperature.

In high temperature ageing two difficulties must be explained: the initial increase in resistance continues for only two hours or so at 160°C as compared with about 100 hours for the increase in strength, and the retrogression effect results in an increase of strength after a preliminary softening. Rohner assumed that at elevated temperatures some of the vacant sites could become electrically neutral regions which could no longer be filled by solvent atoms and which would in fact no longer act like vacant sites but more like grain boundaries. Precipitation would thus be able to take place at those points located inside the grains. The neutral regions were assumed to increase the strength similarly to the vacant sites. It would be possible, however, that vacant sites might join up with neutral regions to form enlarged regions which would have less influence on the strength than the sites and regions separately. The effect of neutral regions upon the resistance was considered to be very much less than that of vacant

sites, since the former would be electrically neutral, and this effect would thus be slight, being due only to the decrease in the conducting cross-section area.

By assuming the formation of neutral regions at the higher temperatures of ageing Rohner was able to explain the hardness retrogression effect and also the fact that retrogression was not observed in resistance measurements.

It was thought that the concentration of sites in the alloy would be limited because the mutual attraction between a negatively charged vacant site and an electropositive ion in interstitial space might cause them to associate. At room temperature, where the conditions would be unfavourable for discharge, a dynamic equilibrium would be attained. This would result in a constant concentration of vacant sites, and the supersaturation in the lattice would not be relieved. At higher temperatures, however, the sites would be discharged continuously, so that the whole supersaturation could be relieved.

Rohner concluded that though the existence of vacant sites could not be directly demonstrated, they were to be regarded as the important factors governing hardness and resistance changes *etc* during ageing.

Some of Rohner's assumptions have been criticized as being of doubtful validity. The idea of interstitial diffusion and the significance of the agreement between the derived theory and the experimental results were questioned, and many considered that the theory was not sufficiently substantiated by experimental data.

CURRENT VIEWS AND FUTURE WORK

Some of the main developments in the theory of ageing have now been summarized, and in conclusion there remains to be indicated the present status of the theory.

There is general agreement that the changes observed during ageing are the result of the decomposition of a supersaturated solid solution. Subsidiary effects introduced by the heat treatments employed are generally regarded as being unimportant so far as the mechanism is concerned. Thus stresses generated during quenching may modify the subsequent course of ageing but do not affect the fundamental nature of the process, and should not be taken as the cause of ageing. This is also true for other similar variables.

Opinions, however, differ as to which of the alterations in the atomic structure during the breakdown of the supersaturated solid solution actually give rise to ageing. Most workers have attributed importance to distortion effects produced in the solid solution lattice in the vicinity of localized segregate regions or precipitate particles. Nevertheless it has been pointed out several times, most recently by Rohner, that it may not be the localized distortions which give rise to the ageing changes but rather the alterations which in the parent lattice arise from the formation of the localized segregates or precipitates. This problem should be investigated further since it is obviously important to decide which factors really affect the behaviour of an alloy. It is a problem intimately connected with the atomic factors governing the strength and electrical resistance *etc* of metals. It becomes of importance to know, for instance, whether defects in the lattice such as Rohner's vacant sites have a greater strengthening effect than the localized distortions postulated in other theories. It may be that both play a part, but further elucidation would be very welcome.

There are many points which are unexplained by Rohner's ideas, but even though some of his assumptions are somewhat doubtful this should not be allowed to bias opinion against his theory as, incomplete though it may be, it has raised a point of much importance.

However if one accepts temporarily the more conventional ideas on ageing changes, that is, that they are caused by the direct influence of segregates or precipitates, the problem still arises of relating changes on an atomic scale to changes in properties which are actually measured. For the aluminium-copper alloys a fairly complete picture has been established of the stages leading to the eventual formation of a precipitate particle of $CuAl_2$ *via* the intermediate stages of Guinier-Preston zones and θ'. Nevertheless, complexity has been recently introduced even into this sequence, since Guinier now postulates a type of superlattice structure in the zones prior to the formation of θ', and Gayler has brought forward evidence for the formation of copper or copper rich aggregates at a certain stage of the ageing. The exact nature of the sequence leading to the formation of $CuAl_2$ is still, therefore, not clear. If one turns to systems other than the aluminium-copper, then there is some evidence to indicate that the breakdown of such solid solutions is similar to that in the aluminium-copper alloys. In the aluminium-magnesium alloys Fink and Smith claimed to have shown that during precipitation the equilibrium phase was

preceded by a transition structure which gradually transformed to the equilibrium phase. This paralleled the later stages of the aluminium-copper alloys, but it was not known whether any stage corresponding to the formation of Guinier-Preston zones occurred.

A transition structure has also been identified in the aluminium-silver system by Barrett, Geisler, and Mehl,[61] and streaks on x-ray photographs, similar to those in the aluminium-copper system, were found. It would appear, therefore, that in this system there is a pre-precipitation aggregation process similar to Guinier-Preston zone formation. Streaks have been observed in the aluminium-magnesium[62] system, where it has been concluded that they are due to an alteration in the structure involving new atomic sites and not merely a shifting of atoms to certain of the existing sites. Despite this knowledge, however, relatively little is known about the changes leading to precipitation, and this lack of knowledge is an obstacle to the formulation of any complete theory of ageing. (But see also note on p. 233.)

If the aluminium-copper alloys are considered as being those about which most is known, then of the observed changes in properties formerly classed as anomalous, the majority have been explained. Unfortunately, however, the correlation between atomic structure and measurable properties such as hardness and electrical resistance is poor, and the fundamental reasons for the observed hardness and resistance changes are still obscure. Mott's explanation of resistance changes on the basis of a critical size of segregate or precipitate exerting the maximum diffracting influence upon the electrons seemed at the time the only possible explanation though strains present in the lattice might exert a considerable effect. Mehl and Jetter have also pointed out the possible importance of strain. It still remains to be proved whether a critical segregate or particle size has more or less influence than lattice distortion, and in what proportion. More knowledge is required about the effect of lattice strain on hardness and resistance and about the distinction if any, that can be drawn between lattice strain generated in different ways.

Owing to the difficulty of correlating observable and measurable changes in properties with changes on an atomic scale, it would seem advantageous to follow ageing by the simplest possible methods, enabling observed results to be related accurately to atomic behaviour. Thus mechanical tests such as hardness and tensile strength should

give way to measurements of say the elastic limit, as the latter quantity would seem easier to interpret in terms of atomic behaviour.

Gayler's recent microscopical work has opened up a new method of study which when used in conjunction with electron microscopy should, in skilled hands, be capable of yielding further information.

There are many additional points upon which relatively little work has been carried out and which are still in need of explanation. The extent and the intensity of the changes produced by ageing and the rate at which they are produced are both questions of considerable importance which merit further attention. In some systems small quantities of solute can produce considerable changes in properties, whereas in others comparatively large amounts of solute exert little effect even though the condition of decreasing solid solubility with decreasing temperature may be fulfilled. The effect of small amounts of additional elements upon the ageing of binary alloys can also be quite marked at times. The reasons for such effects are not known with any certainty, though various theories have been advanced. More work upon such topics would therefore be very useful.

The tendency, in the latest ageing investigations, to use very pure binary alloys represents a considerable improvement on some of the early work where the compositions used were not very carefully controlled. Composition would appear to be an important variable, and careful control is essential in order to ensure results of a fundamental and reproducible nature. An element of doubt as to the validity of any particular theory must always arise if the evidence upon which the theory is based has been obtained by several experiments, each of which used a slightly different composition of alloy.

In conclusion one can say that a great deal has been achieved in working towards a general theory of ageing, but much yet remains to be done. The dispute between the precipitation and pre-precipitation theories largely disappeared when it was realized that the argument was mainly a question of terminology. The problem now to be faced is that of the continuous series of changes involved in the breakdown of a supersaturated solid solution. This breakdown must be studied in a variety of ageing systems by means of techniques which are readily related to fundamental atomic behaviour, so that any common behaviour which is found can be integrated into a general theory of ageing.

NOTE ADDED IN PROOF

A recent review by HARDY[63] contains an excellent summary and analysis of the information available on the changes leading to the eventual formation of equilibrium precipitate phase, by the decomposition of supersaturated solid solutions in various systems. He mentions two different interpretations that can be put upon streaks on x-ray photographs, *i.e.* that they are due to localized segregations as has been commonly assumed, or that they are due to actual formation of thin plates of transitional precipitate. He attempts a reconciliation of the two viewpoints in terms of the particular nature of the atoms in different systems.

REFERENCES

[1] WILM, A. *Metallurgie* 8 (1911) 225
[2] —— *Ibid* 8 (1911) 650
[3] MERICA, P. D., WALTENBERG, R. G. and SCOTT, H. *U.S. Bureau of Standards, Science Paper* 347 (1919); *Trans. Amer. Inst. min. met. Engrs* 64 (1921) 41
[4] —— —— and FREEMAN, J. R. *U.S. Bureau of Standards, Science Paper* 337 (1919); *Trans. Amer. Inst. min. met. Engrs* 64 (1921) 3
[5] JEFFRIES, Z. and ARCHER, R. S. *The Science of Metals* New York, 1924, p 390
[6] MERICA, P. D. *Ibid* p 400
[7] KONNO, S. *Science Reports, Tohoku Imp. Univ.* 11 (1922) 269
[8] FRAENKEL, W. and SCHEUER, E. *Z. Metallk.* 14 (1922) 49
[9] HONDA, K. *Science Reports, Tohoku Imp. Univ.* 11 (1922) 19
[10] GAYLER, M. L. V. *J. Inst. Met.* 30 (1923) 139
[11] HANSON, D. and GAYLER, M. L. V. *Ibid* 27 (1922) 267
[12] TANABE, T. *Ibid* 32 (1924) 415
[13] ROSENHAIN, W. Abstracted by Tanabe, *Ibid* 32 (1924) 438
[14] SACHS, G. *Z. Metallk.* 18 (1926) 209
[15] GÖLER, F. V. and SACHS, G. *Metallwirtschaft* 8 (1929) 671
[16] GAYLER, M. L. V. and PRESTON, G. D. *J. Inst. Met.* 41 (1929) 191
[17] GWYER, A. G. C. *Ibid* 42 (1929) 22
[18] GAYLER, M. L. V. *Metallurgist* 6 (1930) 91
[19] ROSENHAIN, W. Abstracted by MEISSNER, K. L. *J. Inst. Met.* 44 (1930) 226
[20] KOKUBA, S. and HONDA, K. *Science Reports, Tohoku Imp. Univ.* 19 (1930) 365; also summary and critical discussion, *Metallurgist* 7 (1931) 30
[21] TAMMANN, G. *Z. Metallk.* 22 (1930) 365
[22] MERICA, P. D. *Trans. Amer. Inst. min. met. Engrs* 99 (1932) 13
[23] HENGSTENBERG, J. and WASSERMANN, G. *Z. Metallk.* 23 (1931) 114
[24] JENKINS, C. H. M. and BUCKNALL, E. H. *J. Inst. Met.* 57 (1935) 141
[25] Discussion of reference 24, p 182
[26] Reply to discussion of reference 24
[27] GAYLER, M. L. V. Reference 24, p 174
[28] ROSENHAIN, W. *Metallurgist* 7 (1931) 83
[29] FINK, W. L. and SMITH, D. W. *Mining and Metallurgy* 16 (1935) 228
[30] WASSERMANN, G. and WEERTS, G. *Metallwirtschaft* 14 (1935) 605
[31] FINK, W. L. and SMITH, D. W. *Trans. Amer. Inst. min. met. Engrs* 122 (1936) 284
[32] GAYLER, M. L. V. *J. Inst. Met.* 60 (1937) 249

[33] MASING, G. and KOCH, L. *Z. Metallk.* 25 (1933) 137, 160
[34] VOLMER, M. and WEBER, A. *Z. phys. Chem.* 119 (1926) 277
[35] MOTT, N. F. Discussion of reference 32, p 267
[36] See, for instance, FINK, W. L. and SMITH, D. W., ALEXANDER, W. O., STOCKDALE, D. Discussion of reference 32
[37] STOCKDALE, D. Discussion of reference 32, p 277
[38] SUTTON, H. and WILLSTROP, J. W. Discussion of reference 32, p 277
[39] COHEN, M. *Trans. Amer. Inst. min. met. Engrs* 124 (1937) 138
[40] FINK, W. L. and SMITH, D. W. *Ibid* 128 (1938) 223
[41] COHEN, M. Discussion of reference 40, p 233
[42] WASSERMANN, G. Discussion of reference 40, p 243
[43] COHEN, M. Discussion to FINK, W. L. and SMITH, D. W. *Trans. Amer. Inst. min. met. Engrs* 124 (1937) 162
[44] —— *Ibid* 133 (1939) 95
[45] GAYLER, M. L. V. Discussion of reference 40, p 107
[46] CALVET, J., JACQUET, P. and GUINIER, A. *J. Inst. Met.* 65 (1939) 121
[47] MEHL, R. F. and JETTER, L. K. *Age Hardening of Metals* Symposium of the American Society for Metals, 1939, p 342
[48] FINK, W. L. and SMITH, D. W. *Metals Technology* (1939)
[49] MERICA, P. D. Discussion of reference 47, p 417
[50] COHEN, M. Discussion of reference 47, p 432
[51] FINK, W. L. and SMITH, D. W. Discussion to reference 47, p 421
[52] PHILLIPS, A. Discussion of reference 47, p 424
[53] Reference 47, p 435
[54] FINK, W. L. and SMITH, D. W. *Trans. Amer. Inst. min. met. Engrs* 124 (1937) 162
[55] GAYLER, M. L. V. and PARKHOUSE, R. *J. Inst. Met.* 66 (1940) 67
[56] —— *Ibid* 66 (1940) 72
[57] PRESTON, G. D. *Proc. roy. Soc. A* 167 (1938) 526; *Phil. Mag.* 26 (1938) 855
[58] GAYLER, M. L. V. *J. Inst. Met.* 72 (1946) 243
[59] —— *Ibid* 72 (1946) 543
[60] RÖHNER, F. *Ibid* 73 (1947) 285
[61] BARRETT, C. S., GEISLER, A. H. and MEHL, R. F. *Trans. Amer. Inst. min. met. Engrs* 143 (1941) 134
[62] GEISLER, A. H. Dissertation, Carnegie Institute of Technology, Pittsburgh, Pa., 1942
[63] HARDY, H. K. *J. Inst. Met.* 75 (1949) 707

5

HARDENING RESPONSE OF STEELS

E. H. Bucknall and W. Steven

THERE IS in metallurgy no subject which is at once so old and so new as the hardening of steel by heat treatment. Practice in hardening steels goes back to remote antiquity, but only within the closely circumscribed limits imposed by the shallow hardening response of plain carbon steels. Until eighty years ago, when the first useful alloy steel was introduced, it is doubtful whether full hardening had ever been achieved throughout the section of any piece of steel much over one inch thick. Since the introduction of alloy steels both the practice and theory of hardening have advanced rapidly, and it is to be expected that within a few years a basis will have been established for the rational selection of steels for engineering applications and that full advantage will invariably be taken of their hardening capacity. The following account of recent investigations of the hardening depth and transformation mechanism of carbon and alloy steels gives encouragement for this view.

It is well known that the hardening of steels arises from the breakdown of an interstitial solid solution formed by carbon in gamma iron (austenite), the breakdown being associated with the passage of the iron to alpha iron (ferrite) in which carbon is much less soluble than in gamma iron. The iron-carbon phase diagram as usually drawn indicates that in pure iron-carbon alloys the maximum carbon content of austenite is 1·7 per cent by weight at 1130°C, but it is unusual to harden high carbon (hyper-eutectoid) steels from above about 780°C and it is therefore rare for the carbon level of austenite to approach closely to this maximum. The solubility of carbon in austenite falls steadily from 1·7 per cent at 1130°C to 0·8 per cent at 723°C, the eutectoid temperature, with the rejection of iron carbide, Fe_3C (cementite), from austenites of higher carbon contents under normal (metastable) equilibrium conditions. Under the conditions represented by the phase diagram, ferrite separates from hypo-eutectoid steels along a line joining the eutectoid point to the gamma/alpha transition temperature of iron, 910°C. Below the eutectoid temperature the

phase diagram indicates the coexistence of ferrite and cementite, which occur in slowly cooled steels as mixtures of the eutectoid (pearlite) and ferrite or cementite, according to composition. Since cementite is a hard constituent, of approximately 840 D.P.N., decomposition of austenite as indicated by the phase diagram can imply considerable hardening, but it is the ability of austenite to supercool below 723°C and to yield metastable phases in fine states of aggregation or to transform to martensite, which opens up the opportunity of securing great increases of hardness in iron-carbon and related alloys.

Hardening steels to good advantage involves securing a satisfactory approximation to the maximum hardness of which the steel is capable and obtaining this degree of hardening to a required depth, without setting up in the steel either distortion or severe internal stresses likely to lead to immediate or delayed cracking.

Since commercial steels, apart from tool steels and case-hardened steels, are rarely employed in the fully hardened state, a further factor is generally involved, namely tempering.

Tempering tends to obscure evidence of unsatisfactory hardening, and those who handle only finished components are therefore often unaware how common is the failure to secure full hardening in commercial heat treatment. GRIFFITHS, PFEIL, and ALLEN[1] have published photomicrographs of commercial forgings sectioned after oil quenching but before tempering which showed failure to obtain complete transformation to martensite on oil quenching. The structure of one of the components examined by them consisted entirely of a dark-etching intermediate transformation product formed above the martensite range, while others comprised mixtures of martensite and intermediate transformation products, apparently distributed according to a dendritic segregation pattern in the forgings. The areas of intermediate transformation products were associated with inferior mechanical properties, including a lowered fatigue ratio, and are therefore liable to be the areas in which fatigue cracking starts. From the same paper the photomicrographs reproduced in *Plate XXXIII* are taken, showing the structures developed in oil quenched bars, $1\frac{1}{8}$ inch in diameter, of four steels, any of which might be thought suitable for use in that section. As the depth hardening characteristics of these steels improve (from right to left in the illustration), success in achieving a fully martensitic, light-etching, structure is progressively approached, but

even in the left hand picture a small percentage of dark-etching intermediate transformation product is still present. This observation accords with OLIVER's view in his résumé of the activities of the Hardenability Sub-committee,[2] that the largest bar diameter of many of the B.S. 970 En steels which can be fully hardened on oil quenching is surprisingly small and considerably less than the (maximum) diameters in which the steels are recommended for service.

HARDENING IN PLAIN CARBON AND ALLOY STEELS

When dealing with steels other than certain high carbon types which are extremely liable to quench cracking when martensitically hardened and which therefore may be deliberately 'austempered', *i.e.* subjected to isothermal transformation in the lower intermediate ranges, or otherwise only partially hardened, it is the intention when hardening to cool the steel from the austenitizing temperature in such a way that it transforms completely to martensite. This constituent is hard and rather brittle, but on tempering its toughness increases more rapidly than its strength falls off, and it shows, when tempered, better combinations of mechanical properties than higher temperature breakdown products of austenite. The transformation of austenite to martensite occurs over a range of temperatures. In iron-carbon alloys the temperature at which martensite starts to form (M_s) is about 460°C at 0·1 per cent carbon, 330°C at 0·5 per cent carbon, and 160°C at 1 per cent carbon,[3, 91–92] and there is a corresponding depression with increasing carbon contents of the temperature at which the reaction ends (M_f) which is below room temperature at the higher carbon levels, so that sub-zero treatments are necessary if retained austenite is to be avoided.

Rates of cooling intermediate between those which are responsible for pearlitic structures and for martensite give rise to the intermediate transformation products already mentioned. These are not so hard as martensite, and except when formed at low temperatures in highly alloyed steels have lower toughness than martensite when tempered to a given hardness. *Figure 1* indicates the general dependence of the hardness of non-martensitic transformation products on their temperature of formation and on carbon content. The results quoted[4] refer to a wide range of steels. The hardness of the transformation products increases with falling transformation temperature and rises,

rather slowly, with carbon content. There is considerable scatter of results at the high transformation temperatures due to compositional factors, but this becomes small at low transformation temperatures.

The lattice cell of martensite is tetragonal and intermediate in dimensions between the body-centred tetragonal cell which can be used to represent the lattice of austenite and the body-centred cubic cell of ferrite. The centre and corners of the cell are occupied by iron atoms, and the mid-points of the long edges and the centres of the square faces all provide possible sites for carbon atoms, but a maximum of only one twelfth of these sites can be occupied in view of the limited maximum carbon content of austenite of 1·7 per cent.[5] The tetragonality of martensite is evidently caused by the carbon in solution. There is direct evidence in the composition range above 0·6 per cent carbon that the tetragonal ratio is proportional to the carbon content.[6] Rather surprisingly, the hardness of martensite is not directly related to the tetragonal ratio. *Figure 2* shows that hardness increases rapidly at first with carbon content but beyond about 0·6 per cent is practically constant.[7,93–95] The form of relationship raises doubts as to whether the structures concerned are in reality a continuous series of martensites, but the flattening of the curve is unlikely to be due to retained austenite, since treatment below zero does little to alter the position. WRAZEJ[8] has recently expressed the view that austenite is not a uniform solid solution and that only one of the three sub-phases of which it consists breaks down to form martensite. Wrazej's sub-phases are gamma G, which is carbon free iron, gamma S, which has the eutectoid composition and transforms readily

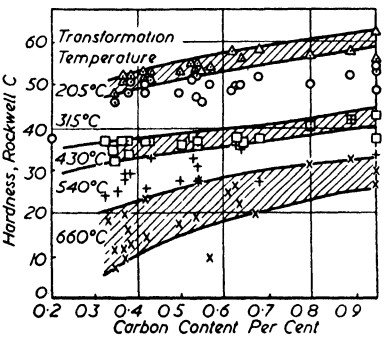


Figure 1. Hardness of isothermal transformation products in relation to temperature of formation and carbon content

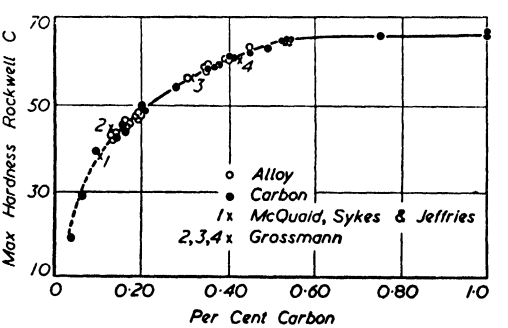


Figure 2. Hardness of fully hardened steels in relation to carbon content

to martensite on quenching, and gamma E, which contains 1·7 per cent carbon and is retained as austenite. In his view the eutectoid composition is 0·888 per cent carbon and contains exactly one carbon atom per twelve unit cells of austenite.[9]

Special attention is drawn to the high values of hardness which may be secured in appropriate sections at low carbon levels in plain carbon steels. These are equivalent to tensile strengths of 80 tons/in^2 (126 kg/mm^2) at 0·1 per cent carbon, 110 tons/in^2 (173 kg/mm^2) at 0·2 per cent carbon, and 130 tons/in^2 (205 kg/mm^2) at 0·3 per cent carbon, rising to 160 tons/in^2 (252 kg/mm^2) at 1 per cent carbon. The hardness figures quoted in *Figure 2* can only be attained if the steel is fully austenitized and all the carbides are in solution. Where carbides remain undissolved at the austenitizing temperature, the effective carbon content of the austenite is, of course, correspondingly reduced.

All commercial steels, whether rated as alloy or plain carbon types, contain other elements in addition to iron and carbon. WEVER[10] showed, as long ago as 1931, that the action of added elements is dependent on their position in the periodic table and on their atomic radius. For instance, the alkali and alkaline earth metals, which have large atomic radii, are insoluble in iron, while the elements with small atomic radii give binary diagrams with iron having open gamma fields (*i.e.* they promote austenite formation), and the transition elements together with certain others give gamma loop diagrams (*i.e.* they promote ferrite formation). The width of the gamma loop is 11·6 per cent for chromium, 4·8 per cent for tungsten, 3 per cent for molybdenum, 2·5 per cent for silicon, 1·1 per cent for vanadium, about 1 per cent each for aluminium and titanium, and 0·55 per cent for phosphorus. When the content of gamma loop elements is progressively increased an increasing proportion of the structure becomes occupied by stable ferrite, which is not convertible to austenite on heating and therefore affords no possibility of transformation hardening on cooling. A similar inability to transform can be induced by addition of austenite promoters, *e.g.* nickel, manganese, nitrogen, copper, and cobalt. Carbon is often listed with the austenite stabilizers, but its main effect is indirect, arising from its strong tendency to combine with ferrite promoting elements.

Figure 3 shows the effects of alloying elements on the maximum solubility of carbon in austenite. All elements lower the solubility, the

ferrite promoting elements acting strongly. As pointed out earlier, lower carbon content of the austenite in hardenable steels is usual in practice, so that the eutectoid carbon content may prove a more important figure than the maximum solubility limit for austenite. At one time it was believed that the eutectoid composition was reduced by all alloying elements, but it is now established[11] that the carbon content of the eutectoid is raised steadily by those elements which form stable single carbides, while other elements, including molybdenum, chromium, tungsten, manganese, and nickel, at first lower but afterwards raise the figure. It is also well known that alloying alters the temperature of the eutectoid formation, which usually occurs over a range of temperature in alloy steels.

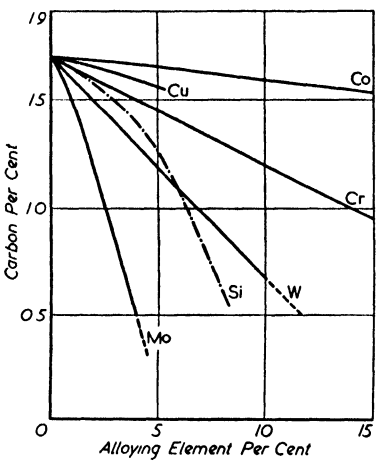


Figure 3. Effects of alloying elements on maximum solubility of carbon in austenite

An extended discussion on the interaction of alloying elements on the constitution of steels at austenitizing temperatures is given in a recent theoretical paper by ZENER.[12]

While the considerations outlined above place limits on the composition and temperature ranges in which alloy steels are capable of full austenitization and of transformation to ferritic structures, within wide ranges alloying elements play only minor parts as regards maximum attainable hardness, which when carbides are fully in solution remains essentially as represented in *Figure 2*, and there is surprisingly little difference between the effects produced by the austenite- and ferrite-promoting elements on the ease with which full hardening is realized. Practically all additions of either type are favourable.

In view of the complex connection between composition and austenitizing behaviour in alloy steels, a practical test to confirm that full hardening can be secured from a selected austenitizing temperature is a necessary stage in laying down the hardening treatment for a new steel.

THERMAL ARRESTS AND COOLING SCHEDULES

The modern outlook on the hardening of steels proceeded from the observation that the critical points on thermal cooling curves for

steels are discontinuously lowered and split by increasing the rate of cooling. The thermal arrests associated with the transformation of austenite on cooling are:

- Ar_3, equilibrium separation of ferrite from hypo-eutectoid steels
- Ar_{cm}, equilibrium separation of cementite from hyper-eutectoid steels
- Ar_1, eutectoid point
- Ar′, depressed eutectoid reaction
- Ar″, intermediate transformation, not usually seen in plain carbon steels
- Ar‴, martensite point.

The designation of the intermediate transformation by Ar″, though common practice, is misleading since several stages of reaction and an extended temperature range may be involved. *Figure 4* is a reproduction of inverse rate cooling curves[13] for a 3 per cent nickel chromium steel cooled at three different, slow rates. At the slowest rate, represented by the right hand curve, the steel suffered a small Ar′ arrest but reached the point *K* substantially untransformed, at which stage a thermal arrest started which apparently reached its maximum intensity at 485°C, though transformation was still proceeding at a much lower temperature, as evidenced by the bulge at 395°C. When cooling was faster, though still relatively slow, the main arrest was concentrated round a lower temperature (467°C on the left hand curve) and the second bulge was eliminated. It is emphasized that all the arrests referred to correspond to intermediate transformations in the steel and reveal the complexity of Ar″, which emerges as a transformation without a fixed temperature of onset for a given steel.

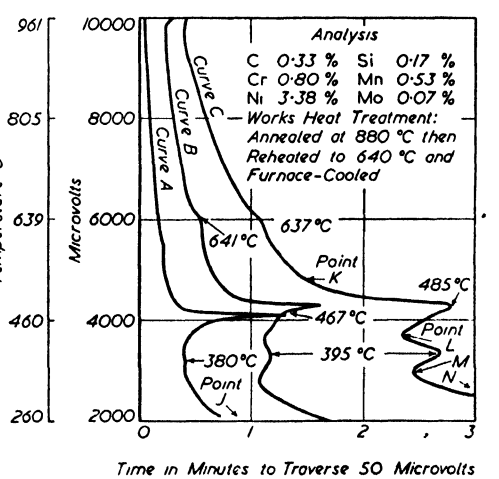


Figure 4. Inverse rate cooling curves for 3 per cent nickel-chromium steel

Cooling curves, such as those given in *Figure 4*, or

more particularly curves extended to embrace a wider range of cooling rates, give information on the critical rate of cooling which must be imposed upon a steel to secure full hardening and on the manner of breakdown of the austenite if the critical cooling rate is not attained, but there are other more convenient ways of studying these matters which yield at least semi-quantitative data concerning on the one hand depth of hardening or hardenability and on the other hand the transformation characteristics of steels. These methods of attack are described later.

It is accepted as an axiom by steel metallurgists that the same cooling schedule when applied to the same steel invariably produces the same microstructure. For a given steel the achievement of a desired structure is entirely a matter of cooling schedules. Plain carbon steels, particularly low carbon steels, require high rates of cooling to yield a martensitic structure. Even with drastic quenching media such as salt solutions unalloyed steels cannot be fully hardened in sections much over one inch in diameter. Additions of alloying elements reduce the critical cooling velocity and allow large sections to be satisfactorily hardened in less drastic quenching media such as oil or even air, which are preferable to water as quenching media because they reduce the danger of cracking and distortion.

The cooling rate developed during a quench is determined primarily by the diameter of the bar being quenched and by the quenching medium adopted, but the rate of flow of the quenching medium and other factors, such as the initial temperatures of the steel and of the quenching medium, the total quantity of quenching medium, the composition of the steel, and the nature of the scale on its surface, may all exert significant effects, as does also the length of the bar unless this exceeds approximately seven times its diameter or special arrangements have been made to restrict end effects.

A widely used index of hardenability is the critical diameter D_c. Strictly speaking, this is the diameter of the largest bar which when quenched in the specified way will develop a fully hardened structure at the axis, but there has grown up a practice of using this phrase to define the bar size giving a half-hard or fifty per cent martensite structure at the axis. The critical diameter is greater for a steel quenched in water than for a steel quenched in oil, and for vigorously stirred oil than still oil. Some of the difficulty which has been experienced in

rationalizing the subject of quenching arises from the latter point, because it proves in practice very difficult to achieve exactly the same effective degree of stirring and efficiency of heat exchange when bars of different diameters are quenched.

Attempts have been made to assess the influence of quenching conditions by the introduction of the concept of quenching intensity. This parameter is based on the simple theory of heat flow which has been presented by RUSSELL[14] and involves assumptions that Newton's law of cooling is obeyed during quenching and that the active thermal characteristics of the steel remain constant over the significant temperature range.

Newton's law of cooling states that the quantity of heat leaving unit surface area of a body in unit time is proportional to the difference of temperature between the surface and the surroundings. If dQ/dt is the rate of loss of heat per unit surface area from a body during the quench and T_1 and T_2 are respectively the instantaneous temperatures of the surface and of the quenching medium:

$$dQ/dt = C(T_1 - T_2) \quad \ldots . \quad (1)$$

where C is a constant.

Further, if dT/dr is the temperature gradient in the body at its surface, then it may be assumed that this is connected with the rate of heat loss from the surface by:

$$dQ/dt = k \,.\, dT/dr \quad \ldots . \quad (2)$$

where k is the thermal conductivity of the steel.

If equations 1 and 2 are combined, we have:

$$dT/dr = C/k\,(T_1 - T_2)$$

The factor C/k, which has the dimension length^{-1}, has been given the symbol h and is used as an index of the cooling power of the quenching medium. In America H is used, where $H = h/2$. h provides a means of conversion from one set of conditions to another, but on its own does not define cooling rate. Thus, excluding stirring effects, the same h value is attributed to a given quenching medium independent of bar size or initial temperature. Recent work by JONES and PUMPHREY[15] indicates that the theory must be used with caution, since calculations from observed cooling rates for oil and water quenching did not accord with constancy of h if the thermal diffusivity of steel was accepted as constant.

This is perhaps not unexpected, as three stages of quenching are recognized for volatile liquids:

1 A stage in which the bar is enveloped in a quiescent vapour film and heat transfer takes place by radiation.

2 A stage in which boiling of the medium occurs and heat transfer is at a maximum.

3 A stage of continuous liquid/solid contact during which heat transfer takes place by convection and where factors such as the viscosity and thermal conductivity of the medium are therefore important.

The relative duration of the three stages depends on the quenching medium. It is commonly urged that a new quenching medium which combined faster cooling in stage 1, so as to limit danger of pearlitic transformation in the steel, and slower cooling in stage 3, so as to reduce stresses in the martensite transformation, would be advantageous.

In order to express hardenability in general terms, critical diameters determined in one medium must be restated in terms applicable to other media. A convenient abstraction is the ideal critical diameter, D_I, which is defined as the critical diameter for an idealized and hypothetical quench which reduces the surface temperature of the bar instantaneously to the temperature of the quenching medium, *i.e.* a quench with h equal to infinity. If for all media the inherent differences in cooling rate could be exactly counterbalanced by changes in bar diameter, generalization of critical diameters could be effected without question. In fact it happens that exactly similar cooling curves cannot be obtained in bars of different sizes quenched in different media, so that a degree of approximation must be introduced. Two widely used criteria of cooling are rate of cooling at 1300°F (704°C) and half-temperature time,[20] which is the time taken to cool to the arithmetical mean of the hardening temperature and the temperature of the quenching medium. In the authors' opinion the former is the more satisfactory.

On occasion it is required to determine the degree of hardening to be expected for a steel of known critical diameter in a section for which the cooling curve is unknown. Provided the thermal diffusivity of the steel can be assumed normal or is known, cooling curves can be calculated to meet such cases. The most usual procedure is *via* the simple theory of heat flow already outlined,[14] but an empirical method described by FRENCH and KLOPSCH[16] is also satisfactory. For simple

shapes, the cooling time taken for the centre to reach a given temperature, minus a correction known as the 'time lag', proves to be proportional to the surface area per unit volume raised to a power n, having a value between 1 and 2. The precise value of n depends on degree of stirring *etc* but average figures are 1·7 for water, 1·4 for oil, and 1·1 for air. The time lag is defined as the time required for two per cent of the total temperature change and is again controlled by the volume/surface ratio. Recent work by POST, FETZER, and FENSTERMACHER[17] has confirmed the factor of French and Klopsch for air cooling.

The development of the differential analyser[18] has opened up the possibility of taking account in cooling curve calculations of variations of the thermal characteristics of steels with temperature. A simple concept arising from such work[19] is that the time lag between the cooling or heating curves at the centre and outside of a bar equals $r^2/4d$ for a cylinder of radius r, where d is the thermal diffusivity. The values of d for steel usually lie between 25 and 50 sq in/hr (161 and 323 cm²/hr). This simple rule applies only for steady cooling and cannot be used in the immediate vicinity of transformation temperatures. Nevertheless it is of great service in estimating temperature gradients within cooling bars.

EARLY MEASUREMENTS OF HARDENABILITY

In GROSSMANN'S pioneer work[20] a series of round bars of a plain carbon steel from $\frac{11}{16}$ in to $2\frac{1}{2}$ in. (1·7 cm to 6·3 cm) in diameter was quenched in water, sectioned, and etched to reveal the hardened case and soft core. In bars having diameters of $1\frac{1}{8}$ in (2·8 cm) or more an unhardened core was found, the size of which increased progressively with bar diameter. When the diameter of the unhardened core, D_u, was plotted against bar diameter D, the diameter at which D_u was zero was readily interpolated.

On other occasions, especially with alloy steels, a sharp line of demarcation could not be brought out by etching, and here Grossmann relied on the indication of the transition from case to core afforded by hardness measurements, interpreted in terms of the hardness associated with a fifty per cent martensite structure $H_{50\%M}$. GROSSMANN[21] has suggested that the half-hard structure has a hardness which is primarily determined by carbon content and is only slightly affected by alloying.

To determine the hardenability on this basis, a series of bars of graded

sizes is quenched, sectioned transversely, and the hardness determined at intervals across the diameter to provide hardness traverse curves. The value of D_u/D is then assessed for each of several curves intersected by a common hardness line, preferably $H_{50\%M}$. Characteristic curves of D_u/D against D can then be drawn and the critical diameter (D_c) derived. Since the steepness of the characteristic curves was shown on Russell's theory of heat flow to be typical of quenching intensity, being steeper for lower values of h, a nomograph can be developed[20] which allows both D_c and h to be determined and D_c to be converted to D_I, the ideal critical diameter.

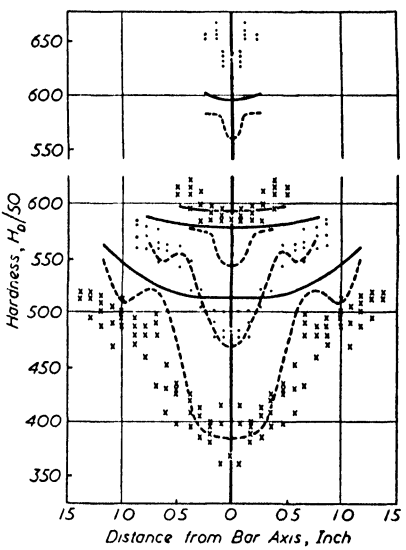


Figure 5. Experimental hardness traverse data and calculated hardness traverse curves

Although direct determination of hardenability on quenched bars may at first sight appear more certain as a method of assessing hardenability than less direct methods, this is not in fact the position. As will have been seen, it is usually necessary to introduce some indirect stage in defining critical diameter, normally hardness measurement. Hardness traverse curves are rarely if ever of the smooth U shape which might be expected. They are usually irregular, and individual results show wide scatter. A typical set of experimental results is given in *Figure 5*,[37] which is discussed in greater detail later. While, with rigorous precautions in quenching, some scatter might have been eliminated, it has been established[34] that transverse variations of hardenability exist within normal commercial steel bars.

Even under uniform quenching conditions smooth hardness traverse curves cannot be obtained, and correlation of one quenching condition with another is difficult. Nevertheless, the study of the hardness traverse curves of quenched bars has proved useful in the study of the response of steels to quenching since the experimental techniques employed relate directly to conditions encountered in heat treatment practice.

THE JOMINY END QUENCH TEST

The study of hardenability by normal quenching methods is both time consuming and laborious, involving the use of considerable quantities of steel in the form of bars of different diameters not always available under laboratory conditions. A simple test which is free from these criticisms was described in 1938 by JOMINY and BOEGEHOLD.[22] The Jominy or end quench test, as it has come to be known, makes use of a single test piece which consists of a cylinder 4 in (9·16 cm) long by 1 in (2·54 cm) diameter with a flange or collar at one end which is used to support the specimen in a quenching jig. The test piece is heated to the normal quenching temperature for the steel under test, held at that temperature for a specified time, and transferred to the quenching jig, which allows the specimen to hang vertically and symmetrically over a ½ in (1·2 cm) diameter water jet orifice as shown in *Figure 6*. The water supply is then turned on and a jet with a free height of 2½ in (6·3 cm) impinges on the end face of the test piece. The steel cools very rapidly at the quenched face and progressively less rapidly towards the flanged end. When cold, the test piece is removed from the jig and its hardness determined at intervals along its length on diametrically opposite flats ground to a depth of 0·015 in (0·38 mm). Specifications for the test have been laid down by the S.A.E.[23] and A.S.T.M.,[24] while the British procedure developed by the Hardenability Sub-committee of the Special and Alloy Steels Committee is described in the recent hardenability symposium.[90]

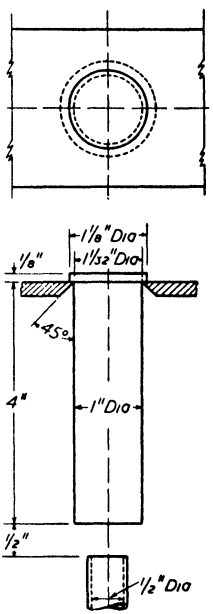


Figure 6. Quenching arrangements in Jominy end quench test

The hardenability curve is obtained by plotting the hardness values against the distance from the end quenched face. A shallow hardening steel hardens only near the quenched face, giving a curve which shows rapid decrease of hardness at a short distance from the end quenched face, similar to curve **A** in *Figure 7*. A deep hardening steel, on the other hand, hardens over the whole length of the test piece, giving an almost horizontal curve similar to curve **B** in *Figure 7*. The standard Jominy test discriminates most effectively between steels in which the main fall of hardness takes place around one inch from the quenched end.

It is less satisfactory for shallow hardening or very deep hardening steels.

Since the rate of cooling of a Jominy test piece during end quenching decreases steadily along the length from the quenched end to the flanged end, the hardenability curve is virtually a curve relating hardness to rate of cooling. The charts issued by the S.A.E. incorporate cooling rate data.[23] The rates of cooling along the length of the test piece have recently been redetermined by RUSSELL and WILLIAMSON.[25] Their results, which did not agree with those previously accepted in America, have been confirmed by a recent investigation.[26] Russell and Williamson's work shows that all the more important low alloy engineering steels cool at similar rates when end quenched, but it should not be assumed that the same cooling rates apply to all steels. Certain steels, *e.g.* silicon-chromium and silicon-molybdenum steels, have significantly different thermal properties which lead to appreciably different cooling rates during end quenching.

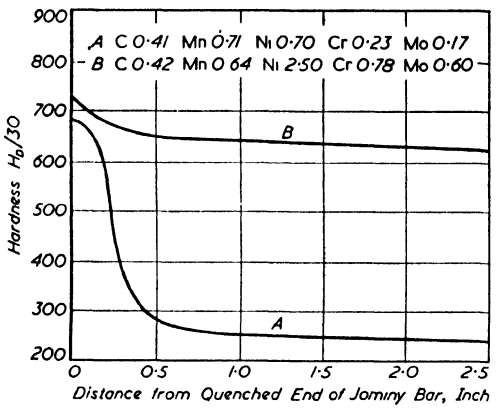


Figure 7. End quench hardenability curves for two steels, A shallow hardening and B deep hardening

It has been shown that the Jominy end quench test is insensitive to quite large variations of end quenching conditions, and no difficulty is experienced in getting satisfactorily reproducible results provided the material tested is of uniform hardenability.[27]

The distance from the quenched end to the position of some specified hardness value, *e.g.* the hardness corresponding to a fifty per cent martensite structure, or to the point of inflection of the curve, often provides a useful index of hardenability, but in America the use of the Jominy test as a routine production control test for successive casts of steel has been mainly in terms of hardenability bands, the whole curve being required to fall within specified limits.[28] *Figure 8* shows the spread of experimental results found in the end quench test as normally carried out on $1\frac{1}{8}$ in bars of En16, a typical En steel.[65] A summary

of the average hardenability curves for representative En steels, as assessed by CRAFTS and LAMONT,[29] is given later in *Figure 12*. An instance where it is essential to report the complete Jominy curve in order to demonstrate a compositional effect has been presented by PUMPHREY and JONES,[30] who have shown that differences in prior heat treatment alter characteristically the shape of the later part of the

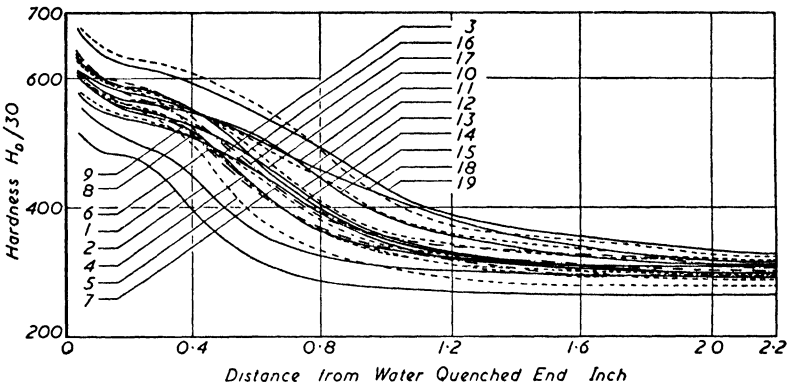


Figure 8. Spread of experimental results for nineteen casts of En16 steel

end quench curve of hyper-eutectoid chromium-molybdenum steels. Corresponding work on titanium steels has been described by NORTHCOTT and MCLEAN.[31]

To convert the results of an end quench test to ideal critical diameters, use is made of a method similar to that adopted for the conversion of D_c diameters to D_I. If the point on the end quenched test piece at which a fifty per cent martensitic structure is developed cools at the same rate as the centre of a bar of diameter D_c when quenched in the specified medium, then D_c is the critical diameter of the steel. The approximations most frequently used to define cooling rates are again the rate of cooling at 1300°F (704°C) and the half-temperature time. RUSSELL[25] has provided a nomograph for such conversions on the basis of half-temperature time.

Figure 9 shows the cooling curves of several points along a length of a test piece during end quenching[25] together with the cooling curves determined at the centres of a series of constructional alloy steel bars quenched in oil. The oil quenching curves have a distinct time lag* before

* This is, of course, the same feature as is recognized in French and Klopsch's work.[16]

the temperature starts to fall, whereas there is no corresponding time lag on the Jominy cooling curves.

If the half-temperature time is used to match cooling curves this difference in shape is given great importance, though the only differences which are likely to be significant are those at possible transformation temperatures, *i.e.* differences in rate of cooling at the lower temperatures. If a Jominy cooling curve and an oil quenching curve have equal half-temperature times they cannot have similar rates of cooling in the transformation ranges.

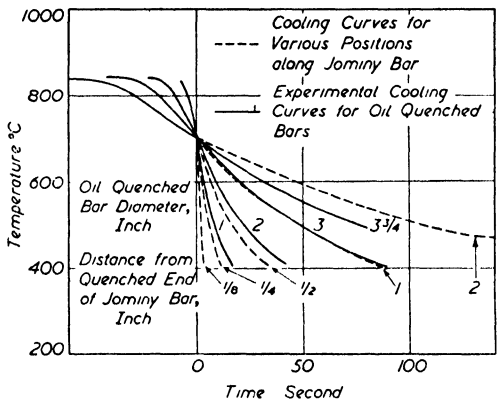


Figure 9. Cooling curves at positions on a Jominy test specimen and at the centres of bars quenched in oil

In *Figure 9* the cooling curves are displaced along the time scale so that all the curves intersect at 700°C. This reveals a marked similarity in shape between Jominy cooling and oil quenching curves below 700°C, particularly in the case of the one inch Jominy position and the cooling curve for the centre of a three inch diameter bar. It would appear that any criterion based on cooling at or below 700°C is to be preferred to one which involves higher temperatures. The cooling curves given in *Figure 9* endorse a point made in the discussion at the hardenability symposium, that the connection between distance along the Jominy bar and diameter of the oil quenched bar with a similar cooling rate at the centre is practically linear over the size range in which the test is most effective, though the connection with ideal critical diameter is less simple. Whatever the system used in conversion, it turns out that a distance along the Jominy bar corresponds to a diameter approximately 2·5 to 3 times as great.

Jominy claimed that the end quench hardenability curve could be used to forecast the hardness distribution within bars, quenched by normal total immersion methods, by following substantially the same procedure as described for the derivation of critical diameters. Experience has shown, however, that the derived results are no more than rough approximations.

The main reason for this is the existence of longitudinal and transverse variations of hardenability within the section of steel bars regarded as commercially satisfactory. If an ingot of steel is rolled to a bar of approximately the diameter of an end quench test piece and hardenability curves are determined on a number of pieces, very little variation of hardenability is observed. Tests on similar samples from different ingots of the same cast in general also show satisfactory agreement. These points are both illustrated by *Figure 10*, which shows the hardenability variation observed in $1\frac{1}{8}$ in (2·8 cm) diameter bars derived from the top, middle, and bottom of the first, tenth, and eighteenth ingots of a cast of En35 steel.[33] In these tests only the surface material has been tested. If, on the other hand, end quench test pieces are machined from off-centre positions of a larger bar, considerable variation of hardenability can readily be demonstrated. For example, when test pieces from positions in a billet as indicated in *Plate XXXIV* were end quenched in the normal manner and flats ground for hardness testing, wide variations in hardenability were revealed,[34] as shown in *Figure 11*. Considerable evidence was accumulated by the Hardenability Sub-committee to show that longitudinal and transverse variations of hardenability were characteristic features of normal commercial steel. These variations do not follow a strict pattern, but certain broad trends have been observed. The hardenability of the steel near a bar surface is usually fairly uniform, but the hardenability at the axis varies longitudinally and is usually lower than that near the surface, although

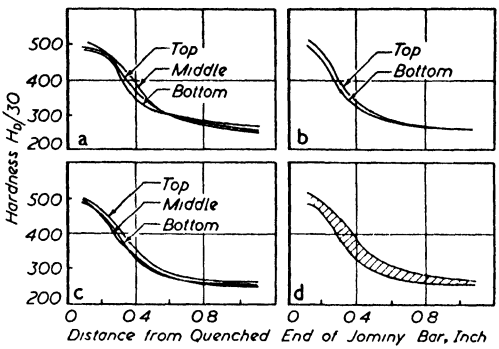


Figure 10. Hardenability curves for three ingots from one cast of an En35 steel. a *Ingot 1* b *Ingot 10* c *Ingot 18* d *Total range of hardenability observed in ingots 1, 10 and 18*

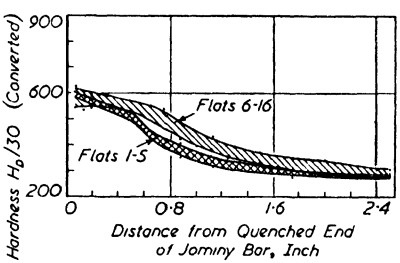


Figure 11. End quench hardenability curves for different positions within section of NE8630 steel

examples have been reported where it is higher. There is also frequently a peak hardenability somewhere between the surface and the axis, but the magnitude and position of this may vary along the bar.[32,37]

A method has been described for revealing the pattern of the hardenability variation across the section of a bar, involving macro-etching transverse slices which have been incompletely isothermally transformed at a temperature where the material of higher hardenability undergoes slower transformation than the material of lower hardenability.[34] If transformation is arrested by quenching at an opportune stage, a structure will be developed consisting substantially of intermediate product or pearlite in the zones of low hardenability and largely of martensite in the zones of high hardenability. A section treated in this manner is shown in *Plate XXXIV.*

Additional tests made by the authors have emphasized the frequent occurrence of hardenability variations across the section of billets. For instance, in a series of thirteen billets of En steels a distinct variation in hardening response which could be revealed either by the Jominy test or by etching after partial intermediate transformation was found in seven billets, in all but one of which the steel in the centre of the billet had lower hardenability than the steel near the surface. The variation was specially strong in the billet of En17 (manganese-molybdenum steel), for which at one inch from the quenched end of the Jominy bars the Rockwell C hardness varied from 46 to 36 according to position within the billet section. The results of this investigation are summarized below:

Table I. Hardenability Variations within thirteen B.S. En Steel Billets

Macro-etching		*Hardenability Testing*	
Result	*No. of Billets*	*Result*	*No. of Billets*
Uniform	4	*Uniform*	4
Nearly uniform	2	*Almost uniform*	2
Faint square	2	*Soft centre*	3
Distinct square	3	*Very soft centre*	3
Prominent square	2	*Slightly hard centre* ..	1

Chemical analyses carried out near the outside and the centre of the billets showed only very minor differences in composition which were no greater for the steels with marked variation of hardenability than for those which were substantially uniform.

The cause of such hardenability variations is not altogether clear. Some may be accounted for by chemical segregation on a macroscale, but others are not readily explicable. For example, the hardenability pattern shown in *Plate XXXIV* cannot be accounted for by chemical segregation or grain size variation. Analytical surveys by normal methods and by the traversing spectrograph technique both failed to yield positive results. BARDGETT has reported that oxygen segregation does not account for the variations, but so far as is known the possible influence of hydrogen has not been examined.[35]

The zones of low hardenability frequently observed at the centre of bars may prove to be accounted for ultimately by dendritic or micro-segregation. The fact that banding is usually more marked towards the axis of a bar is evidence of a greater degree of microsegregation at this position. Segregation on this scale would not be revealed either by normal analytical methods or by current spectrographic techniques. It is conceivable that the zones of lower alloy content, and therefore of lower hardenability, would act as nuclei promoting the breakdown of the zones of higher alloy content, giving an overall effect of decreased hardenability relative to the more uniform surface material. End quench hardenability curves frequently indicate that the central zones of low hardenability are not capable of developing as high a maximum hardness as zones of high hardenability near the bar surface. This might possibly be due to formation of less soluble carbides in the highly alloyed zones of microsegregation or alternatively to a greater tendency for the retention of austenite.

In view of the known variations of hardenability within steel bars, the hardenability of a cast of steel or even of a bar of steel cannot be defined adequately by a single end quench hardenability curve but only by a band of curves determined at different positions within the steel. The most satisfactory, simple method so far available of obtaining such a representative set of curves involves end quenching a test piece machined from a position half an inch off centre from a $2\frac{1}{4}$ in to $2\frac{1}{2}$ in (5·7 to 6·3 cm) diameter bar and testing the hardness on three flats ground 90° apart.[34] There is every likelihood that a series of tests of

this type would give a fair indication of the total range of hardenability existing within the product of a cast of steel. If then all subsequent considerations were based on the curve of lowest hardenability, a satisfactory margin of safety would be automatically ensured. A similar though rather more elaborate procedure has been recommended by WALKER, ECKEL, HINO, and MUELLER.[36]

The known transverse and longitudinal hardenability variations largely account for the failure of the method proposed by Jominy for forecasting quenched bar hardness traverses from end quench hardenability curves. For example, *Figure 5* shows hardness results across En110[37] steel quenched in oil in several different sizes, the experimental points being indicated by dots and crosses, curves calculated from Jominy end quench tests being superimposed. The curves drawn as full lines were calculated using an end quench hardenability curve determined on a sample machined from $1\frac{1}{8}$ in (2·8 cm) diameter bar, thus ignoring transverse hardenability variations. The broken curves were calculated taking into account the hardenability variations. It will be seen that the agreement is much improved as compared with that for the full curves though still by no means exact. The experimental technique involved, however, is extremely cumbersome, and in practice it would be much simpler as well as more accurate whenever hardness traverse data are required to section quenched bars of various sizes than to arrive at indirect estimates of the transverse variation of hardness.

It must be accepted that the end quench test cannot be used to forecast the hardness distribution within quenched bars accurately. Nevertheless, as pointed out by JONES and PUMPHREY,[38] there is no doubt that a relationship exists between the hardness of the Jominy bar and of oil quenched rounds. Indeed, the test would serve no useful purpose as a measure of hardenability if an approximate correlation did not exist. Although experimental scatter is great the correlation is high enough to be useful, so that if a steel is required which will develop a certain hardness in a certain bar size it is possible by considering end quench hardenability curves to sort available steels into three classes:

1 those which are definitely too shallow hardening,
2 those which are marginal,
3 those which have adequate hardenability.

To test the depth hardening potentialities of a wide variety of steels

by normal hardening and tempering methods would involve a great deal of work, but, by using the Jominy test as a preliminary selection test, those steels which are definitely inadequate can be readily eliminated. Modern specifications refer to maximum ruling sections of 6, 4, 2½, 1½, and 1⅛ in (15·2, 10·2, 6·4, 3·8, and 2·9 cm), of which the two extremes fall outside the range within which the standard Jominy test can be usefully applied. For 2½ and 1½ in diameter sections the specifications demand certain mechanical properties at the centre, while for a 4 in section mid-radius properties are specified. It has been pointed out that for a steel to stand any chance of meeting specification requirements in 1⅛, 2½, and 4 in oil hardened sections it must develop adequate hardness $\frac{7}{16}$, $\frac{15}{16}$, or 1⅜ in (11·1, 24·6, or 34·9 mm) along the end quench test piece, while further steels can be eliminated if a certain excess of hardness is demanded as a tempering allowance, the amount depending on specification requirements, *e.g.* it may be anything from 50 D.P.N. upwards.[32]

Perhaps the most striking example of the usefulness of the end quench test was its successful application in the development of the American NE (National Emergency) steels. In 1943 the Americans were faced with the necessity of severely reducing their consumption of certain alloying elements. This demanded the development of an entirely new series of steels to replace the standard S.A.E. steels, which would make full use of the alloying elements in more plentiful supply and make the least possible demand on those metals in short supply. In view of the limited time available the more orthodox methods of developing new steels were impracticable, and the end quench test had to be used as the chief method of assessment of interchangeability. The principle adopted was that if the end quench hardenability of one of the new NE steels was similar to that of an S.A.E. steel, then it was regarded as a possible substitute. Its ultimate acceptance as an official replacement depended on the results of subsequent experience in production and application.

In this connection it is interesting to recall that CRAFTS and LAMONT [29] demonstrated the existence of a systematic relationship between the hardenability and the mechanical properties required of the B.S. En series of steels. They classified the average end quench hardenability curves of the various En types into groups according to the tensile strength specified at a maximum ruling section of four inches. The

average curve for each of these groups is reproduced in *Figure 12*. It will be noted that the end quench hardenability and the specified tensile strength increase uniformly up to a tensile strength of 60 tons/in^2 (94 kg/mm^2).

TESTS FOR SHALLOW HARDENING STEELS

Other tests have to be applied in studying hardenability outside the Jominy range. The Shepherd PF (penetration fracture) test, although one of the oldest hardenability tests, is still the most satisfactory for shallow hardening steels of the low alloy tool steel type. For this test a series of samples, each $\frac{3}{4}$ in diameter by approximately 3 in long, is quenched from different austenitizing temperatures under standard conditions in 10 per cent brine. After quenching, the samples are fractured transversely near their mid-length and the fracture examined to determine the influence of quenching temperature on the grain size. The surfaces of the fractured faces are then ground and etched to determine the depth to which the steel has hardened. The Shepherd hardenability, or P factor, of the steel is given by the depth of hardening, expressed in sixty-fourths of an inch.[39]

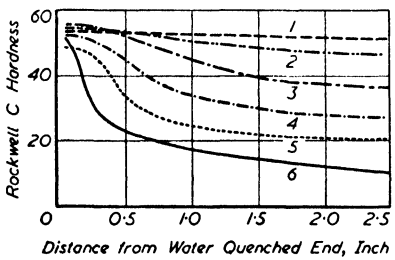


Figure 12. Composite diagram of Jominy curves representing the average degree of hardenability in En steels in terms of the tensile range for use at 4 in maximum ruling section. 1 60/100 ton min. (En 23 and 25–30) 2 55/60 ton (En 17 and 24) 3 50/60 ton (En 16, 19, 22 and 100) 4 45/55 ton (En 18 and 21) 5 40/50 ton (En 13, 14 and 15) 6 Carbon (En 5, 8 and 9)

The transition line from hardened outside to unhardened core, indicated satisfactorily by etching the steels of the type normally subjected to this test, corresponds fairly closely to a structure containing 50 per cent martensite and 50 per cent pearlite.

The results of a PF test are not normally used to calculate the depth of case that would be obtained with other sizes of bar or other quenching media, but are used to allocate casts to the specific applications which on the basis of previous experience appear to be appropriate at the hardenability levels observed.

The Shepherd disk test, like the PF test, is primarily intended for shallow hardening tool steels.[40] A series of disks, each at least $1\frac{3}{4}$ in (4·4 cm) in diameter and varying in thickness by $\frac{1}{32}$ in (0·79 mm), is hardened in brine under standard conditions. After hardening each

disk is cut in half, polished, and etched. The hardenability of the steel is indicated by the thickness of the thinnest disk which shows signs of a soft core when etched.

A number of tests for shallow hardening steels which use modifications of the Jominy technique have been recommended in the American literature,[23,24] but these are not much used in this country.

TESTS FOR DEEP HARDENING STEELS

The Jominy end quench test cannot be used to assess the hardenability of the deep hardening alloy steels which are normally air hardened, since these harden fully even at the slowest cooling rates obtained on an end quench test specimen. POST, FETZER and FENSTERMACHER[17] have described a test, in principle similar to the end quench test, which they claim is capable of assessing the relative hardenabilities of this type of steel. A specimen of the steel is screwed into the end of a block of steel six inches long by six inches in diameter so that a test length four inches long by one inch diameter extends from the end of the block. The assembly is heated to the normal hardening temperature and then allowed to cool in air. The part of the test length adjacent to the block cools at the slowest rate and points farther along the test length cool at progressively faster rates, up to the cooling rate of a one inch bar cooled in air. Thus a plot of the hardness developed along the length of the sample provides a hardenability curve similar to the end quench hardenability curve but referring, of course, to a different range of cooling rates. An obvious source of error in this test is the possibility of a variation from test to test of the efficiency of thermal contact at the screwed end of the sample, but with reasonable care the test provides a useful extension to the range of the end quench test. The interpretation of the results of course requires just as much caution as in the end quench test.

The S curve provides another basis for systematic study. The transformation/temperature/time diagram of a steel may be used for estimating hardenability if determined from the same austenitizing temperature as used for hardening the steel. If a higher austenitizing temperature is used, transformation in the pearlitic range may be much delayed, though not in the intermediate range. S curves are normally determined by transferring suitably austenitized specimens of the steel to isothermal decomposition baths (usually lead baths) held

at fixed temperatures below Ar_3, the progress of some physical property change closely related to transformation then being traced.[42] Suitable properties are electrical resistance, magnetic saturation intensity, and the dilatation accompanying transformation. The microstructure of specimens quenched from the bath after intermediate periods of exposure can also be used as a basis of S curves.

The connection between S curves and hardenability may not at first sight be apparent. It is in any case inferential. The S curve indicates the ability of a steel to transform above the martensite range, and its resistance to such transformation is measured in hardenability tests. *Plate XXXIII* shows a close connection between maximum rate of transformation and the proportion of intermediate transformation product found in oil quenched $1\frac{1}{8}$ in diameter bars for a series of steels, all of which had their maximum rate of transformation in the intermediate range. The figure used here to express rate of transformation is the maximum transformation velocity, which is defined as the velocity constant of a first order reaction proceeding at the same rate as the observed fastest transformation (when transformation is fully under way).[42] In the steels of *Plate XXXIII* the proportion of intermediate transformation product falls from about 70 per cent to under 10 per cent as the maximum transformation velocity decreases from 10 to 3 parts per minute. Steels having their fastest transformation in the pearlitic range can tolerate higher transformation velocities for a given percentage transformation because of more rapid cooling in the pearlitic range than in the intermediate range.

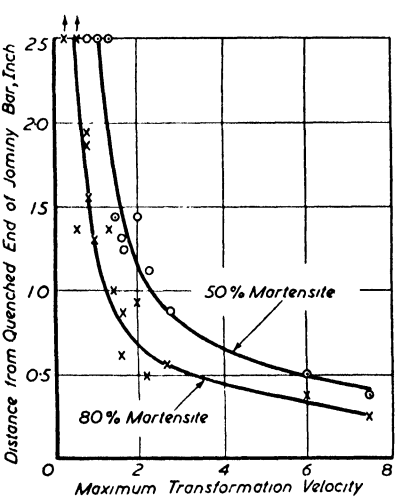


Figure 13. Diagram relating maximum transformation velocity to Jominy end quench hardenability

COOPER and ALLEN[43] have given further data which relate transformation velocity with hardenability as measured in the Jominy test (*Figure 13*). It will be seen that with few exceptions the experimental results fall close to a curve indicating an inverse relationship between the two parameters. The work of Cooper and Allen encourages

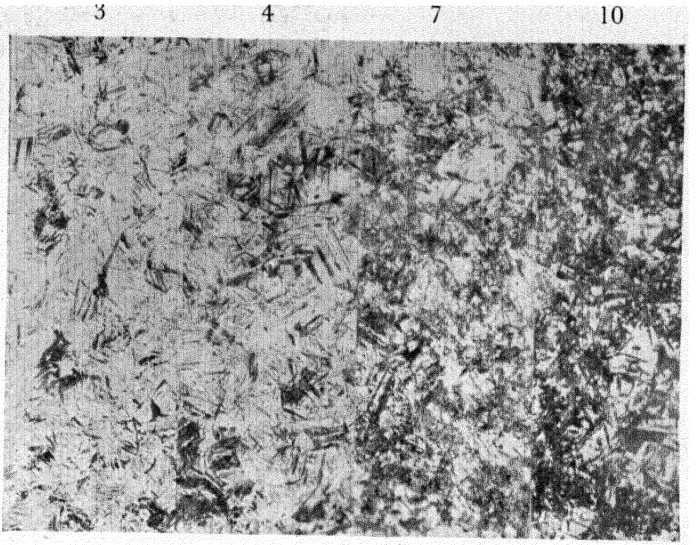


Plate XXXIII. Alloy steels oil quenched in $1\frac{1}{8}$ in diameter bars. Maximum transformation velocities as above the illustration

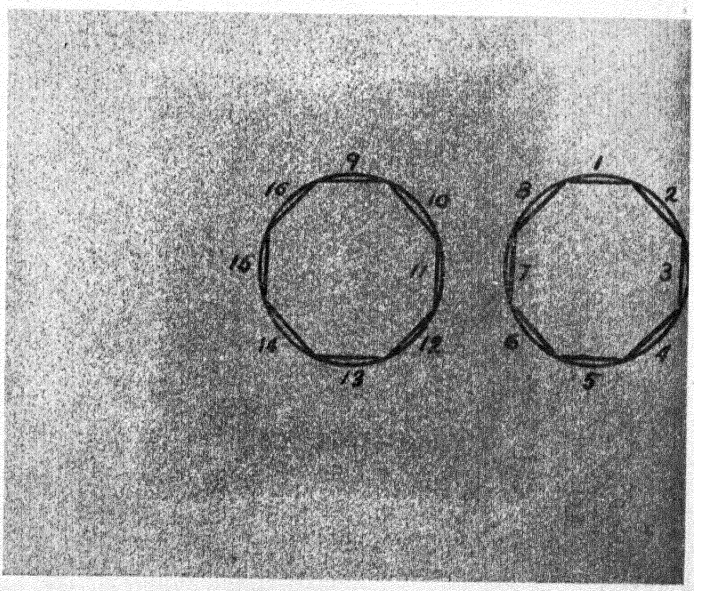


Plate XXXIV. Hardenability pattern revealed by etching NE8630 steel after partial isothermal transformation

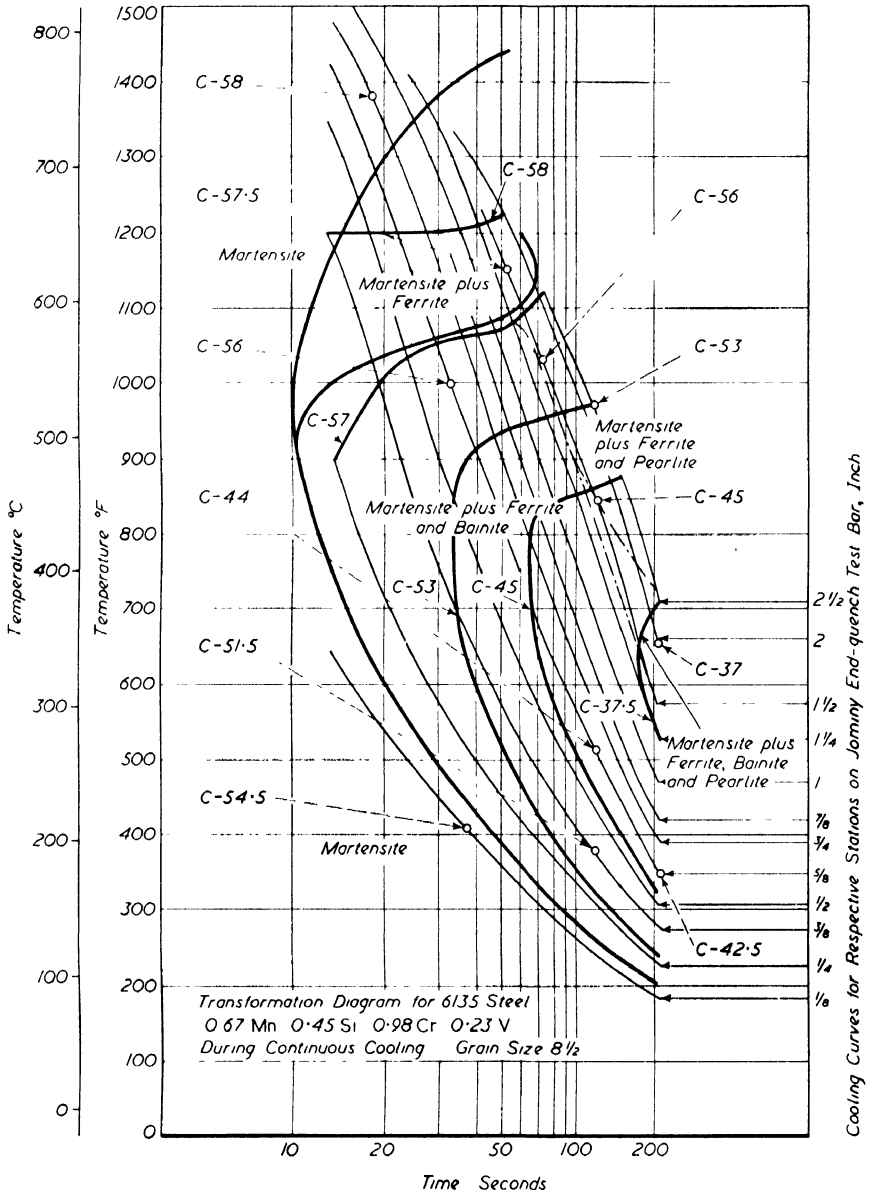


Figure 15. Continuous cooling transformation diagram for S.A.E. steel 6135. Rockwell hardness is indicated by the dotted arrows. (After Liedholm)

the idea that analysis of S curves may give satisfactory information on the hardenability of air hardening steels which are not differentiated by the Jominy test. It is, however, not always feasible to derive from an S curve a single index of the hardenability of a steel.

Figure 14 shows the S curve for an En110, nickel-chromium-molybdenum, steel determined by a combination of dilatometric and microscopic methods. It will be seen that this steel has two temperature ranges of rapid decomposition of austenite, the first around 625°C in the pearlitic range, and the other below 500°C in the intermediate range. In this steel the latter type of transformation is the faster and between the two ranges of rapid transformation there is a temperature zone in which austenite is relatively stable. Diagrams of this general form are typical of steels with fairly high contents of chromium and other gamma loop elements. Higher contents of these elements than occurring in the En110 steel would move back the pearlitic 'nose' and extend the zone of austenite stability without altering the general features of the diagram. Steels of this type would be expected to behave at fast cooling rates approximating to their critical cooling rates as indicated by their transformation characteristics in the intermediate range, transformation above the martensitic range setting in, if at all, when the 500°C shelf is crossed; though for slow rates of cooling, transformation would occur in the pearlitic range. For such a steel, the maximum transformation velocity is clearly not fully descriptive of the depth hardening characteristics.

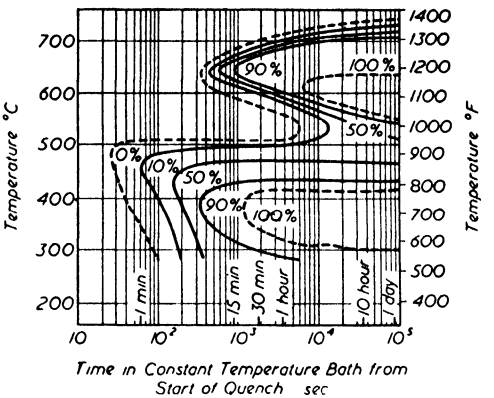


Figure 14. Transformation/time/temperature diagram for En110 steel

TRANSFORMATIONS ON CONTINUOUS COOLING

Another method of approach is the study of continuous cooling diagrams. Only a few of these diagrams are as yet available in the literature and their determination is by no means simple. A continuous cooling transformation diagram[41] such as that reproduced in *Figure 15*

indicates the critical cooling rate which must be exceeded to secure a fully austenitic structure in the steel examined, and in addition shows the course of martensitic breakdown during more protracted cooling.

When S curves first began to appear it was customary for the assumption to be made that they applied exactly to continuous cooling as well as to isothermal conditions and therefore to interpret them by reading off the intersections of a curve denoting the cooling rate. BAIN,[44] however, recognized at an early date that continuous cooling diagrams must be different from the corresponding S curves and expressed the view that the curve denoting start of transformation on the diagram must be an integrated version of that shown on the other. The problem has since been attacked by many workers,[45–50, 82] most of whom have followed the lead of SCHEIL[45] in stating that the fractional times spent in preparation for transformation at different temperatures are approximately additive, at least where they concern a single type of transformation. This generalization has recently been used by RUSSELL[51] when calculating quenching stresses. A different basis of integration was used by GRANGE and KIEFER,[47] who dealt with constant cooling rates and not with normal cooling.

It is evident that the precise cooling schedule rather than the average cooling rate needs to be known before a fair assessment can be made of transformations on cooling. Ideally there is for each steel a different cooling diagram for every type of cooling, *e.g.* for constant rate cooling, cooling by oil quenching, or cooling as in the end quench test.

JAFFE[50] has recently studied non-isothermal transformation in two nickel-chromium-molybdenum steels, the one a hypo-eutectoid steel, and the other a hyper-eutectoid steel. He first determined the start curves for isothermal transformation in terms of the times required for one per cent transformation, as estimated microscopically, and then went on to investigate the effect of holding the steels for varied periods at higher or lower temperatures between quenching and isothermal exposure, the periods chosen for the intermediate stage being in each case only a fraction of those required to reach the one per cent values. He found that pro-eutectoid ferrite and carbide transformations are each approximately additive, but that holding austenite in the ferrite range accelerates the bainite reaction by an amount significantly greater than that to be expected from the additive fractional time principle. An

opposite deviation from additivity applied to the results of exposure at one temperature in the bainite range followed by isothermal transformation at another.

It is a criticism of the end quench test as normally conducted and reported in terms of hardness measurements that no differentiation is made between a large degree of non-martensitic transformation at low temperatures, where the isothermal transformation product is hard, and transformation to a smaller extent at a higher temperature, where the transformation product is softer (see *Figure 1*). Microscopical examination would usually discriminate between such cases, but a readier method of investigation is necessary where a number of temperatures of transformation may be involved. In such cases the most convenient procedure is to carry out a series of end quench tests on bars of the steel under investigation, exposing the bars to the end quench for progressive times and then transferring them bodily into a drastic quenching medium. Hardness determinations along the bars may then be made and interpreted in terms of the temperature attained at the time of quenching, using the known cooling schedules of points along the Jominy bar. Experiments of this sort have been described by PUMPHREY and JONES[52] for a series of hyper-eutectoid steels, while PUMPHREY[53] has made a detailed examination of a 0·8 per cent carbon, 1·1 per cent chromium steel for which the recorded hardness curves on end quenched bars show that up to one inch along the bar transformation has occurred mainly in the interval between 100 and 150 seconds but that a little farther along the bar transformation is still proceeding after longer times. Interpretation of these results in terms of temperature showed that transformation proceeded closely in accordance with the type of S curve normally obtained for this kind of steel. More recently PUMPHREY and JONES[54] have investigated further steels, for each of which they determined the S curve and the end quench curve. The observed hardness of the Jominy bar approximated closely to a calculated hardness value derived by summing fractions of the hardness values of the isothermal transformation products formed in each interval of temperature, the appropriate fractions being derived from the S curve. The steels examined included both hypo- and hyper-eutectoid types, and the results would seem to indicate that although the hypothesis is not exact it gives a reasonable working approximation for a wide range of steels.

In general, better correlation would be expected at the initial stages of a continuous cooling diagram and the corresponding section derived mathematically from the S curve than between corresponding curves for later stages, since at most temperatures the first precipitation is of a different composition from the original austenite and causes alterations in base composition, often increasing both the carbon and alloy contents. In extreme cases this may give a composition to which the transformation diagram for the original steel is quite inapplicable.

EFFECTS OF INCOMPLETE HARDENING ON MECHANICAL PROPERTIES

In the recent symposium on hardenability CRAFTS and LAMONT[29] expressed a view long held in this country when they emphasized the need for the full martensitic hardening of steels in ensuring the optimum mechanical properties after tempering. The need becomes particularly pressing for steel components which are highly stressed in service, particularly if stressing occurs at sub-zero temperatures, is multi-axial or involves shock. In the laboratory the disadvantages of non-martensitic transformation have been demonstrated repeatedly by mechanical tests on tempered steel bars, which have been:

1 subjected to stepped quenching, with partial or complete transformation at an intermediate holding temperature,[1,43,55,56]

2 deliberately cooled in such a way as to cause some transformation above M_s.[57] For example, CRAFTS and LAMONT[29] used the ingenious idea of taking Izod impact bars from Jominy test pieces, locating the notches at predetermined points with known cooling characteristics equivalent to $2\frac{1}{2}$ and 4 inch diameter bars, *viz* $1\frac{1}{8}$ and $1\frac{13}{16}$ inches from the quenched end,

3 unsatisfactorily hardened by normal quenching.[43,58]

The test results were compared in each case with those given by fully hardened bars, tempered either to the same strength or at the same temperature as the partially hardened bars.

Method 1 may appear the least direct but in practice it is the most informative, since the maximum effects of transformation at a variety of temperatures, as well as the effects of partial transformation, can readily be studied in this way.

Figure 16 is a summary of a great number of mechanical tests on a variety of constructional steels,[43] each steel being tested in two basic conditions, oil quenched in test bar size and fully transformed at the temperature of fastest transformation in the intermediate range, followed in each case by appropriate tempering. The diagram shows that it is sometimes possible at the lower tensile levels for a good combination of tensile strength and impact value to be reached by intermediate transformation products, but the risk of falling below specification requirements is great and increases with tensile level. Similarly elongation, reduction of area, and proof stress deteriorate when non-martensitic transformation occurs.

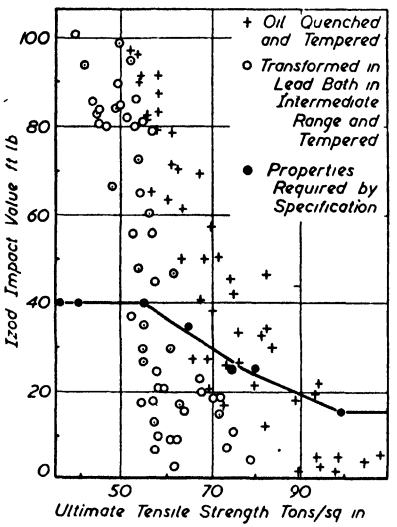


Figure 16. Izod impact values plotted against tensile strength for fully hardened and isothermally transformed steels

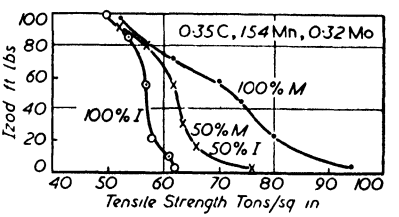


Figure 17. Effect of partial and complete isothermal transformation on Izod/tensile relationship of a 1·5 per cent manganese-molybdenum steel

The stated deficiencies are those which appear in steels tempered to equal tensile strengths; when tempered at equal temperatures a deficiency in tensile strength appears and the shortcomings in the other properties are correspondingly reduced. The extent of the deterioration proves to be greater the lower the alloy content, the higher the transformation temperature above the martensite range, and the lower the tempering temperature, with exceptions in notch impact value within the 'Izod trough'.

Partial intermediate transformation causes a degree of deterioration roughly proportional to the amount of the transformation product and to its effect when occupying the whole structure. Morphology as well as constitution enters into the question, rounded masses having much smaller effects in general than needles. Attention has therefore

to be paid to the multiple forms taken by non-martensitic reaction products.[84] The effect of partial and complete transformation on a 1·5 per cent manganese-molybdenum steel is shown in *Figure 17*. This steel is more seriously affected than most 0·35 per cent carbon steels; high carbon steels are liable to rather more serious deterioration.

The deterioration of properties caused by non-martensitic transformation is not confined to the results of the normal mechanical tests. GRIFFITHS, PFEIL and ALLEN[1] have shown corresponding effects on torsion and fatigue characteristics. *Table II* summarizes their results on the fatigue strength of steels in the fully hardened and tempered condition with that of the steels isothermally treated at either 400° or 450°C so as to effect partial transformation, and then tempered to a hardness equal to that of the hardened and tempered steel.

Table II. Fatigue Strength of some Hardened and Tempered Steels

Percentage Composition	*Treatment* *L.Q. = lead quenched* *O.Q. = oil quenched* *T. = tempered*	*Brinell Hardness* (2/120/30)	*Endurance Limit* 10^7 *Reversals* (*tons/in²*)
C 0·38 Cr 1·11 Mo 0·42	*L.Q.* 830–400°C *for* 30 *min.*, *T.* 620°C *O.Q.* 830°C, *T.* 660°C	248 253	± 27·5 ± 31·7
C 0·37 Ni 1·25 Cr 0·56	*L.Q.* 830–450°C *for* 30 *min.*, *T.* 580°C *O.Q.* 830°C, *T.* 650°C	229 240	± 25·5 ± 28·7
C 0·41 Ni 1·96 Mo 0·31	*L.Q.* 830–450°C *for* 40 *min.*, *T.* 580°C *O.Q.* 830°C, *T.* 630°C	251 252	± 27·9 ± 30·0
C 0·4 Ni 3·5	*L.Q.* 830–450°C *for* 10 *min.*, *T.* 500°C *O.Q.* 830°C, *T.* 600°C	260 268	± 30·5 ± 32·7

Partial intermediate transformation in each case reduced the endurance ratio from approximately 0·55 to 0·50. Similar results have recently been presented by HOLLOMON, JAFFE, MCCARTHY, and NORTON,[56] who studied the effect of both pearlitic and intermediate transformation and also showed that the endurance limit is not reduced by temper brittleness.

Figure 18 shows these authors' results for the effect of non-martensitic transformation on the Charpy impact value of an NE8735 steel

containing 0·34 per cent carbon, 0·66 per cent nickel, 0·56 per cent chromium, 0·24 per cent molybdenum. The reduction of impact value at and above room temperature is greater for pearlitic than intermediate transformation, but both types of partial transformation cause the onset of low temperature embrittlement around 0°C. The embrittling effect of non-martensitic products increases rapidly with decreasing temperature of test.

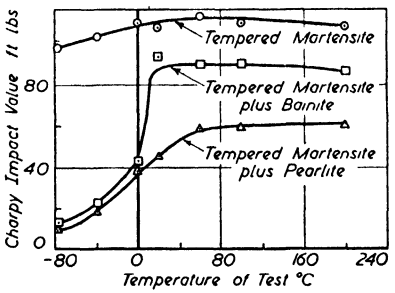


Figure 18. Impact properties of NE8735 steel at tensile strength of 55 tons/sq in in relation to partial pearlitic and intermediate transformation

Figure 19 shows the variations in mechanical properties associated with different centre hardness values, expressed as abscissae in the diagrams, of an En17 manganese-molybdenum steel tempered to three fixed levels of tensile strength, namely 64, 59, and 56 ton/in² (101, 93, and 88 kg/mm²). The data are taken from HONEYMAN,[57] who used an air blast to produce the required centre hardness in seven steels. The properties studied were tensile strength, M.S., r ratio of yield point to tensile strength, Y.R., and Izod impact value at 18°C and −40°C. The general results are similar to those reported for stepped quenching, and the low temperature notch impact test again proved extremely sensitive to as-quenched hardness.

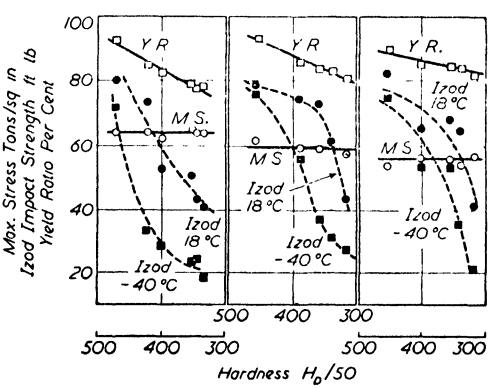


Figure 19. Mechanical properties of En17 steel tempered to 64, 59, and 56 tons/sq in after hardening to varied extents as shown

Much information on the mechanical properties which result from quenching steels in sections below and above the maximum at which they will harden fully is found in the *Report of the Technical Advisory Committee on Special and Alloy Steels*, B.S. 971, which provided the background for the bold step taken by the committee in laying down maximum sections in which the steels were to be used at given tensile levels.[58] *Figure 20* is drawn on the basis of data taken from this source,

for an En16 manganese-molybdenum steel, and shows the tempering characteristics of this steel when hardened in different sections to give different degrees of hardening.[32] The full, uppermost curve is theoretical, representing the behaviour expected if the steel was fully hardened to give the characteristic hardness shown in *Figure 2*. The lower slopes of the tempering curves for the partially hardened steel are largely the result of tempering being restricted to the martensitic part of the structure. The impact values for this steel proved to depend much more on tempering temperature than on success of hardening, but the impact value at a given tensile strength was, of course, much impaired by incomplete hardening.

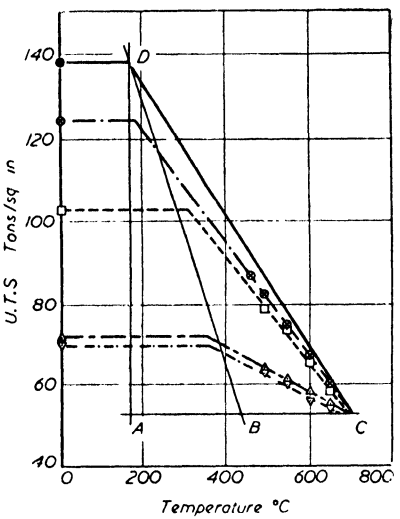


Figure 20. Tempering of En16 steel after hardening to different extents, based on B.S. 971 data

Mechanical property variations have been studied for eleven En steels by Cooper and Allen,[43] who have discussed the general connection between the hardenability of these steels and their ability to meet rigorous aircraft specifications which call for mechanical properties at the centre of bars in excess of those required at the mid-radius position by B.S. 970. The steels were heat treated at the maximum ruling sections for 55, 65, 80, and 100 ton (87, 102, 126, and 157 kg/mm^2) specifications. At the 55 ton/in^2 tensile level it was found that there was a close correlation between the hardenability and the capacity to meet the specification, but at the 65 ton level exceptions begin to appear, which continued at the 80 ton level and became the rule at the 100 ton level, where only a selection of the steels having the adequate hardenability met specification requirements. They concluded that although hardenability is essential it is not the only factor, and that factors concerned with toughness become increasingly significant at high tensile levels. In Cooper and Allen's tests many criteria of hardenability were considered. A still more pronounced departure from complete correlation between hardenability and capacity to meet specifications would presumably have occurred if

the normal 50 per cent martensite criterion alone had been used. Incidentally, this paper directs attention to the unequal demands made by specifications as regards accompanying properties at the different tensile levels. *Figure 16* is particularly interesting in this connection. The requirements of a 100 ton steel are, for example, much nearer the maximum values of which present day steels are capable at this level than the properties required in 60 ton steels, where the test requirements can be met by a completely non-martensitic steel, though this might not give a good account of itself in service without a high factor of safety.

In concluding this aspect of the subject, a few words seem necessary on austempering, a deliberate isothermal transformation in the lower intermediate range, and martempering, *i.e.* ordinary quenching delayed so as to reduce temperature gradients to a minimum before passing M_s. There seems to be a common belief that the austempering of high carbon steels gives them mechanical properties in excess of the maxima to be achieved by full hardening and tempering. This only occurs where micro- and macroquench cracks have reduced considerably the 'fully hardened' properties, and in such cases still better properties are to be obtained by martempering, which reduces the risk of cracking to a minimum. Some alloy steels benefit by martempering, especially those intended for application at tensile levels above 100 ton/in^2.

It should also be mentioned that there are special applications of alloy steels for which non-martensitic transformation is advantageous. One is in steels to be welded, where cracking of the heat affected zones alongside the welds is to be feared whenever these harden above 350 D.P.N.[59] Another is in ferritic steels for prolonged service in the temperature range of 800 to 1000°F, where tempered high temperature intermediate products give better creep resistance than tempered martensite.[60] A third concerns the matrix of hardenable cast irons of the acicular iron class, where the composition is designed so that on cooling the casting a particular variety of intermediate transformation product is produced together with as small a proportion of martensite as possible.[61]

ATTEMPTS TO SYSTEMATIZE HARDENABILITY DATA

When a series of steels varying progressively in their contents of one or more alloying elements is quenched in a size which gives partial hardening, it is usually found that there is a definite relationship

governing the hardness obtained. For example, when the effect of additions of chromium and molybdenum to a 0·9 per cent nickel steel was investigated by the Jominy test,[62] it was found that the results could be expressed simply in terms of the amount of molybdenum required together with fixed percentages of chromium to give definite levels of hardenability.

GROSSMANN[20] made such observations with a series of carbon and alloy steels which were quenched in a series of diameters permitting estimation of the critical diameter for a 50 per cent martensitic structure at the centre. The results encouraged him to make the first important attempt to rationalize the influence of grain size and alloying elements on the hardenability of steel. He suggested a method of calculating the hardenability of steels from grain size and chemical composition in terms of ideal critical diameter, D_I. For this purpose a base hardenability is first allocated to the steel according to its carbon content and grain size and this is multiplied in turn by factors for each of the alloying elements. The base hardenability increases with carbon content and coarseness of grain size, though not steeply. For instance, for a straight iron-carbon alloy with 0·20 per cent carbon and a fine A.S.T.M. grain size of 8 Grossmann derives an ideal critical diameter of 0·14 in (3·6 mm). Increase of carbon content to 0·90 per cent increases the ideal critical diameter to 0·29 in (7·4 mm) at the same grain size or to 0·40 in (10·2 mm) at a coarser grain size of 4. The chart from which base hardenability is read is reproduced in *Figure 21*. Turning to the alloying elements, Grossmann's work showed that all alloying elements except sulphur and cobalt are beneficial to hardenability. Some elements, including molybdenum, manganese, and chromium (when properly taken into solution), were found to have large effects, while silicon, nickel, and copper appeared to have smaller though still definite hardening powers.

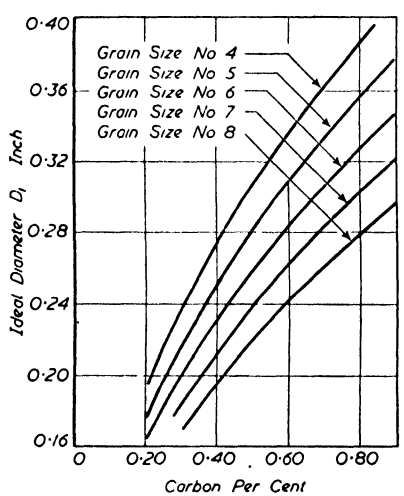


Figure 21. 'Base-hardenability' diagram, showing effects of carbon content and grain size

Grossmann's estimates of the roles played by several of the common

elements are given in *Figure 22*, where the effectiveness of the element is expressed as a factor by which the base hardenability must be multiplied to take account of the amount of alloy present. The uncertainty of the chromium factor for chromium percentages exceeding 0·4 is due to the possibility of undissolved carbides.

According to Grossmann, increase in content of each individual element has an additive effect but for combinations of alloying elements the individual factors have to be multiplied. For example, 0 per cent, 1 per cent, and 2 per cent manganese give factors of 1, (1 + 5·4), and (1 + 2 × 5·4) *i.e.* 1, 6·4, and 11·8. When present together 1 per cent manganese and 1 per cent chromium, which individually have factors of 6·4 and 2·8, are credited with a joint factor of 6·4 × 2·8, *i.e.* 17·9 and not 6·4 + 2·8, *i.e.* 9·2.

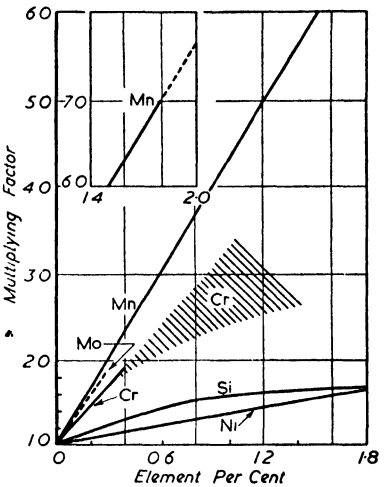


Figure 22. Multiplying factors for alloying elements

The computation of Grossmann's ideal critical diameter from composition and grain size is greatly facilitated by use of published nomographs.[63,64] The simplicity of the Grossmann method is attractive and adequate accuracy is obtained for some steels,[65] though general experience has shown that the results are frequently erroneous, especially for steels of medium or high hardenability.[62,66] Grossmann attached considerable significance to the additive effect of progressive additions of a given element in contrast with the multiplying effect for several elements. He used this as an argument for encouraging the use of a number of small additions of alloying elements rather than a larger quantity of a single addition. The present authors consider that it has become difficult to maintain this argument in favour of multi-alloy steels, though certain combinations of alloying elements have undoubted advantages as will be seen when the position is discussed from the S curve standpoint. Reassessment of Grossmann's factors on the basis of Jominy tests have been made by CRAFTS and LAMONT,[67] HODGE and OREHOSKI,[67] KRAMER, HAFNER, and TOLEMAN,[68] KRAMER, SIEGEL, and BROOKS,[69] and by BROPHY and MILLER,[70] all of whom

indicate frequent non-linearity of the relationship between alloy content and effect on hardenability, *e.g.* an upward trend of the characteristic curves for manganese and nickel, while maximum effects have been shown for other elements when present at intermediate levels. The characteristic curves of nickel in low alloy steels as determined by Brophy and Miller are reproduced in *Figure 23*. It will be seen that the effectiveness of the added nickel generally increases progressively with the nickel content of the steel. Between 3 and 5 per cent nickel has a marked effect on hardenability, though in the range 0 to 2 per cent with which earlier workers were concerned its contribution as measured in the Jominy test is small. Crafts and Lamont have extended their calculations to include not only the computation of hardenability but the prediction of the hardness[71] and Izod impact values[72] of heat treated steels. Comstock[73] has suggested that for titanium additions hardenability correlates not with total content but with that part of the titanium content which is acid soluble, while Grange and Garvey[74] have taken a similar view that in boron treated steels the boron can be present in effective and ineffective forms. Boron in small amounts of the order of 0·003 per cent can greatly improve hardenability.

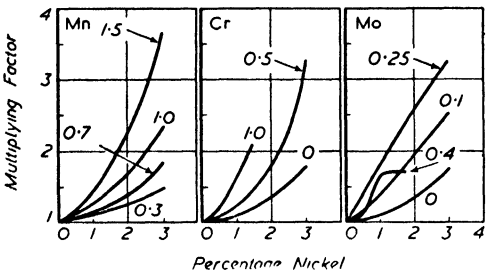


Figure 23. Multiplying factors for nickel in presence of manganese, chromium and molybdenum the percentages of which are indicated by figures on the curves

In Grossmann's original work a 50 per cent martensite criterion of hardening was employed, as discussed earlier. In constructional alloy steels the 50 per cent martensite point does not define accurately the change from the fully hardened to the normalized condition, and in any case it is below the degree of hardening required for high tensile applications. Other criteria of hardenability more suitable for alloy steels, *e.g.* 99 per cent or 95 per cent martensite, have been considered by some investigators.[87] These criteria have the disadvantage that they are difficult to apply accurately to the interpretation of Jominy curves, and microscopic examination must be introduced to supplement hardness measurement. According to Hollomon and Jaffe,[76] ideal critical diameters for fully hard and half-hard structures invariably

have a ratio near 0·75 : 1, but GRANGE and GARVEY[74] consider that although this ratio may apply to carbon steels it does not apply to low alloy steels.

A weakness of the Grossmann method is the assumption that an element makes the same contribution to hardenability irrespective of the nature and amount of other alloying elements present. The work of some of the American authors referred to above suggests that some elements become more effective when present in combination than when alone, but GLEN[75] has taken the opposite stand. Glen examined some 400 high frequency furnace melts which covered variation of manganese, variation of nickel and manganese together, either in the presence or absence of molybdenum, variation of manganese and chromium together, variation of molybdenum when present with nickel or with chromium, and variation of molybdenum and manganese together when present with nickel or with nickel and chromium. His results confirmed recent American work indicating that the effect of successive increments of a given element is not constant, but showed that each alloying element is most effective when no other addition is present, the presence of second and third alloying elements causing a decrease in the effect of the first element to an extent depending on the elements concerned. In general, Glen's results are not surprising, but there is one aspect which is not easy to explain. Additions such as manganese and nickel have been shown to give concave factor curves, that is to say when they are added alone their effectiveness increases according to the basic hardenability of the steel to which they are added, and it is difficult to see why this should not apply when they are present in combination.

Jominy tests indicate that almost all alloying elements increase the hardenability of steels. Definite exceptions are sulphur, which acts by abstracting manganese from the austenite, and cobalt,[77] while aluminium has an indirect effect in the same direction due to its action on grain size. Elements such as vanadium which readily form carbides may act in a way analogous to sulphur when low austenitizing temperatures are used, by lowering the carbon content and hence the hardenability of the austenite. The mechanism of the action of cobalt is still obscure. SMOLUCHOWSKI[78] has reported that, unlike most other addition elements, cobalt increases the diffusion coefficient of carbon

in austenite, which might promote transformation in both the pearlitic and intermediate ranges, but HAWKES and MEHL[79] have found no evidence for this nor anything to support the possibility that cobalt acts by an indirect mechanism, such as that of aluminium and vanadium. Their work established, however, the fact that cobalt decreases the incubation period and increases the rate of transformation at all temperatures between the eutectoid point and the martensite range.

In the course of measurements of the transformation characteristics of alloy steels it was found that the relationship between maximum transformation velocity and alloy content for steels could often be expressed at least over limited ranges by a multiplying rule, equal increments of alloy content reducing the maximum transformation velocity in a way which is constant for a given alloying element. The experimental results show a linear relationship when the logarithm of the maximum transformation velocity is plotted against percentage of alloying element. For example, in nickel-chromium steels each addition of 0·1 per cent carbon between 0·1 and 0·6 per cent reduces the transformation velocity to 0·4 of its former value. The corresponding factors for other elements appeared to be:

Table III. Maximum Transformation Velocity Reduction Factors

Element of which 0·1 per cent is added	*Maximum Transformation Velocity Reduction Factor*
Carbon (with nickel and chromium)	0·4
Nickel (with chromium)	0·90
Chromium (with molybdenum)	0·88
Chromium (with nickel)	0·84
Manganese	0·76
Molybdenum (with nickel)	0·68

For steels with similar S curves it is interesting to note that the maximum transformation velocity is approximately inversely proportional to the diameter in which a steel can be fully hardened by oil quenching. The results of the transformation measurements when expressed in this form give similar curves to those found by Brophy and Miller.[70] Further checking against the Brophy and Miller data

showed, however, that no simple basis could be found for systematizing hardenability results for multialloy steels.[80]

Further experimental studies of transformation indicate that the attempt described above to systematize transformation data, and hence to compute hardenability, over-simplifies the position. For example, when the effect of an addition on maximum transformation velocity is studied over a wide range of compositions it is often found that its beneficial effect does not continue indefinitely.

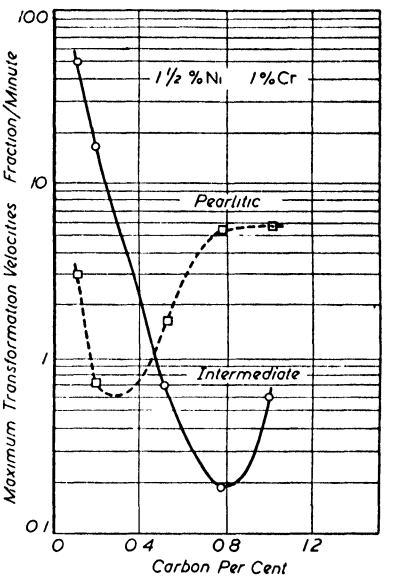


Figure 24. Effect of carbon on maximum transformation velocity of 1¾ per cent nickel-chromium steels

Figure 24, for instance, shows new data on the effect of carbon on the transformation characteristics of 1¾ per cent nickel-chromium steels. It will be seen that while early additions of carbon lower steeply the rate of transformation in both the pearlitic and intermediate ranges, there is an optimum effect on each type of reaction at an intermediate carbon content beyond which further increase in carbon speeds transformation. The optimum proportion of carbon is different for the two reactions, so that while the low carbon steels of the series have their greatest liability to transformation in the intermediate range, the higher carbon steels are most liable to transformation in the pearlitic range. LYMAN and TROIANO[81] have found that in 3 per cent chromium steels increase of carbon content from 0·08 to 1·28 per cent progressively speeds up pearlitic transformation. Another element which does not appear to have a consistent delaying effect on transformation in the pearlitic range is chromium. MANNING and LORIG[82] have given S curves for a series of hypo-eutectoid chromium-molybdenum steels of fixed carbon content, including two with 1·5 per cent and 2·4 per cent chromium. Between these levels, increase in chromium content had very little effect on the higher temperature region of the pearlitic range either as regards the pro-eutectoid stage or the formation of pearlite. Again, the effect on the

beginning of the lower intermediate transformation was small, though the increased chromium content delayed transformation at higher temperatures of the intermediate range. There is, therefore, in general, little reason to suppose that the hardenability of the steel had been improved appreciably by the increased chromium content. Manganese may be a further alloying element without a consistent progressive effect. At first additions have a progressively increasing effect, but this trend seems unlikely to continue to high manganese contents, for POWELL[83] has shown that a 13 per cent manganese, 1·2 per cent carbon steel transforms from the austenitic condition slowly but apparently completely on exposure at about 450 to 500°C, as indicated by a fall in electrical resistance which begins in rather less than ten hours.

It is not uncommon for alloying elements to have specific effects confined to particular stages of transformation. Molybdenum, for example, though having the reputation of affecting powerfully the whole pearlitic transformation, actually does little to suppress the formation of pro-eutectoid ferrite, its action being due almost entirely to post-poning and restricting the progress of the formation of the carbide, of which it is a constituent. This is due, no doubt, as MEHL[84] has suggested, to its slow diffusion through the untransformed austenite. Other work indicates that boron has the complementary effect, causing opposition only to pro-eutectoid ferrite separation.

The simultaneous presence of several alloying elements in a steel may complicate greatly the mechanism and progress of transformation. For example, LORIA[85] has recently shown that in an air hardening nickel-chromium-molybdenum-vanadium steel of practically eutectoid carbon content (0·60 per cent C, 2·75 per cent Ni, 1·25 per cent Cr, 0·50 per cent Mo, 0·12 per cent V) intermediate transformation at 400 to 450°C involves no less than three distinct structural stages. The initial precipitation produces carbon-rich plates, believed to be formed by a shearing process, but this is a transition phase which later decomposes with diffusion of carbon, with rising temperature, resulting first in the nucleation of ferrite grains along the surface of the plates and later in the transformation of the neighbouring austenite, by this time enriched in carbon, to a ferrite/carbide aggregate, this stage proceeding progressively until all the austenite has broken down.

Transformation studies have shown clearly that the hardenability of those steels which transform rapidly in the pearlitic range is much

improved by grain coarsening, because this reduces the grain boundary area at which nuclei of ferrite and cementite appear. No similar effect of grain size is found in the intermediate, bainite, range, where nucleation is not confined to the grain boundaries, so that for compositions which have their fastest transformation in this range no sacrifice of hardenability accompanies selection of fine grained steels with their better optimum mechanical properties. The bainite range is that within which most present day alloy steels, which owe their hardenability largely to carbide formers, transform isothermally most readily, and the question arises whether adequate use is being made of the elements which do not promote bainite formation (carbon, silicon, manganese, nickel, and copper) in heat-treatable steels. If it is recognized that all non-martensitic transformation should be avoided, a logical approach to the problem of manufacture of steels of high hardenability would seem to be first to suppress the possibility of bainite formation by adding elements of the type referred to above, and afterwards to push back the pearlitic 'nose' of the S curve by means of additions of carbide formers, boron *etc.*

Of great interest in this connection is the suggestion of HOLLOMON and JAFFE[76] that hardenability data could be satisfactorily systematized if separate account was taken of the effect of addition elements on the pearlitic and bainitic transformations. According to their view, much of the hardenability data given by previous workers should be assumed to refer to pearlitic hardenability. Grain size was considered to have no direct effect on bainitic hardenability, but changes in carbon content were regarded as having the same effect as in pearlitic hardenability. On the basis of available isothermal transformation data they assumed nickel and manganese to have the same effect on bainitic hardenability as on pearlitic hardenability. Molybdenum, on the other hand, was considered to have no effect at all on bainitic hardenability. As chromium appeared to have some effect on bainitic hardenability, though less than on pearlitic hardenability, it was regarded as having one half the effect measured for pearlitic hardenability.

Though Hollomon and Jaffe's system is only tentative and, as they admit, no direct measurements have yet been made of bainitic hardenability, it is the most promising scheme of systematizing hardenability yet developed because it is less empirical than earlier schemes and bears a nearer relation to the picture developing from transformation

studies. Hollomon and Jaffe tabulate calculated hardenability figures for American commercial steels, and make the comment that while the pearlitic hardenabilities of some of the steels are adequate for heavy sections their bainitic hardenabilities are generally low, particularly where full hardening is made the criterion of assessment.

The following table contains a selection of the calculated figures for the S.A.E. steels which are of most interest in this country:

Table IV. Calculated Hardenability of S.A.E. Steels

SAE Type	*Approximate Composition*	*Calculated Hardenability (Ideal Critical Diameter in Inches)*			
		Pearlitic		*Bainitic*	
		Full-hard	*Half-hard*	*Full-hard*	*Half-hard*
1030	0·3% C	0·7	0·9	0·75	1·4
1330	1·6% Mn, 0·3% C	1·4	1·9	1·5	2·8
2330	3·5% Ni, 0·3% C	1·9	2·5	2·0	3·7
5140	1% Cr, 0·4% C	2·5	3·3	1·8	3·2
3130	1·25% Ni, 0·6% Cr, 0·3% C	2·8	3·7	2·1	3·8
3240	1·75% Ni, 1·1% Cr, 0·4% C	4·2	5·5	2·9	5·2
4340	1·75% Ni, 0·65% Cr, 0·35% Mo, 0·4% C	7·7	10·3	3·1	5·7
4130	1% Cr, 0·2% Mo, 0·3% C	2·8	3·7	1·2	2·2

The view is taken that the Hollomon and Jaffe concept of bainitic and pearlitic hardenability will eventually form the basis of a method of hardenability calculation giving a close approximation to the truth. At present, however, bainitic hardenability factors are not available for some of the important alloying elements and some of those which have been suggested for others cannot be accepted as reliable, for example, the value for chromium appears to be high and that for nickel low. The method of combining the factors must also be reconsidered since the present method of multiplying the factors together, on the assumption that the influence of an element is independent of the nature and amounts of other elements present, is undoubtedly over simplified. No system of hardenability prediction can be exact, however, which does not explain the observation that steels of practically identical compositions, and even parts of the same steel, can differ widely in hardenability.

CONCLUDING REMARKS

The successful hardening of steels depends basically on their equilibrium constitution in the hardening temperature range and on the factors involved in attainable maximum hardness, but in practice by far the most important factor in hardening engineering constructional steels proves to be that termed 'hardenability'. Hardenability is not an exact concept, but this fact does not reduce its practical significance. This is evidently still not sufficiently recognized, since steels are still commonly employed in sections in which they cannot be fully hardened and cannot therefore show their optimum mechanical properties after tempering.

ALLEN[86] has expressed the extreme view 'that there is no such property as hardenability at all. The effect of an alloying element on hardenability is the resultant of the effect of the alloying element on all the reactions which may take place between the quenching temperature and the temperature at which martensite forms, and the effect of the alloying element on the hardness of such reaction products as appear.' While there is much reason in this statement, a wider definition of the field of hardenability research would allow the analysis of hardenability into its component parts and the recognition of the separate origins and individual effects of these components not only on hardness but also on the mechanical properties of steels. In the past there has been too much over-simplification of the subject, which was little advanced by the publication of work in which the merits of those alloying elements which violently oppose pearlitic transformation were emphasized to the detriment of elements acting more gently in this range. These, even when non-uniformly distributed on the dendritic pattern, do not produce sudden transitions from soft to hard zones or severe local stress and may offer more adequate opposition to bainitic transformation. The introduction of the duplex concept of hardenability by Hollomon and Jaffe is a welcome step towards the recognition of the dependence of hardenability on transformation characteristics and points the way to a more rational basis for the employment of alloying elements in constructional steels.

Unsolved problems still remain in this general field, of which the most interesting is the origin of hardenability variations within apparently homogeneous steels. Related problems include the factors influencing quench cracking and the partition of alloying elements between

carbide and ferrite in tempered products. JANITSKY and BAEYERTZ[88] and PATTON[89] have shown that the mechanical properties of fully hardened steels are practically unaffected by composition up to a tensile strength of about 80 ton per square in (126 kg/mm^2) but the influence of composition on the properties of partially hardened steels is still largely unsolved. These problems are, however, receiving close attention and the rate of recent technical progress in the study of the hardening response of steels justifies expectation of an early solution.

The authors acknowledge their indebtedness to The Mond Nickel Company Limited for permission to publish this chapter.

REFERENCES

[1] GRIFFITHS, W. T., PFEIL, L. B. and ALLEN, N. P. *Second Report Alloy Steels Research Committee, Iron and Steel Inst. Special Report No.* 24 (1939) 343
[2] OLIVER, D. A. *Symposium on the Hardenability of Steel. Iron and Steel Inst. Special Report No.* 36 (1946) 417
[3] ZENER, C. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 550
[4] United States Steel Corporation, *Atlas of Isothermal Transformation Diagrams* Pittsburgh, 1943
[5] PETCH, N. J. *J. Iron Steel Inst.* 147 (1943) 221
[6] KURDJUMOV, G. and LYSSAK, L. *Ibid* 156 (1947) 29
[7] BURNS, J. L., MOORE, T. L. and ARCHER, R. S. *Trans. Amer. Soc. Metals* 26 (1938) 1
[8] WRAZEJ, W. J. *Nature, Lond.* 158 (1946) 308
[9] —— *J. Iron Steel Inst.* 152 (1945) 189
[10] WEVER, F. *Mitteilungen aus dem Kaiser-Wilhelm-Institut fur Eisenforschung* 13 (1931) 183
[11] SHAPIRO, C. L. and STRAUSS, J. *Trans. Amer. Inst. min. met. Engrs* 158 (1944) 335
[12] ZENER, C. *Ibid* 167 (1946) 513
[13] JENKINS, C. H. M., BUCKNALL, E. H. and JENKINS, G. C. H. *Iron and Steel Inst. Special Report No.* 24 (1939) 181
[14] RUSSELL, T. F. *First Report Alloy Steels Research Committee, Iron and Steel Inst. Special Report No.* 14 (1936) 149
[15] JONES, F. W. and PUMPHREY, W. I. *J. Iron Steel Inst.* 156 (1947) 37
[16] FRENCH, H. J. and KLOPSCH, O. Z. *U.S. Bureau of Standards Technologic Papers No.* 313, 1926
[17] POST, C. B., FETZER, M. C. and FENSTERMACHER, W. H. *Trans. Amer. Soc. Metals* 35 (1945) 85
[18] BUSH, V. *J. Franklin Inst.* 212 (1931) 447
[19] JACKSON, R., SARJANT, R. J., WAGSTAFF, J. B., EYRES, N. R., HARTREE, D. R. and INGHAM, J. *J. Iron Steel Inst.* 150 (1944) 211
[20] GROSSMANN, M. A., ASIMOW, M. and URBAN, S. F. *The Hardenability of Alloy Steels* New York, 1939, p 124
[21] —— *Trans. Amer. Inst. min. met. Engrs* 150 (1942) 227; *Metal Progress* 42 (1942) 80

[22] JOMINY, W. E. and BOEGEHOLD, A. L. *Trans. Amer. Soc. Metals* 26 (1938) 574
[23] *S.A.E. Handbook* New York, 1944
[24] *A.S.T.M. Book of Standards* 1 (1942) 1106
[25] RUSSELL, T. F. and WILLIAMSON, J. C. *Iron and Steel Inst. Special Report No.* 36 (1946) 34
[26] BOEGEHOLD, A. L. and WEINMAN, E. W. *Office of Scientific Research and Development, Project N.R.C.* 55, No. 2382, 1944
[27] Section IIA *Symposium on the Hardenability of Steel. Iron and Steel Inst. Special Report No.* 36 (1946) 9
[28] See for example, *Steel* (1945) 116
[29] CRAFTS, W. and LAMONT, J. L. *Iron and Steel Inst. Special Report No.* 36 (1946) 283
[30] PUMPHREY, W. I. and JONES, F. W. *Ibid* 400
[31] NORTHCOTT, L. and MCLEAN, D. *J. Iron Steel Inst.* 157 (1947) 492
[32] BUCKNALL, E. H. *Ibid* 157 (1947) 67, 72
[33] Section VD *Iron and Steel Inst. Special Report No.* 36 (1946) 113
[34] BUCKNALL, E. H. *Ibid* 120
[35] BARDGETT, W. E. *J. Iron Steel Inst.* 157 (1947) 65
[36] WALKER, H. L., ECKEL, E. J., HINO, J. and MUELLER, F. H. *Metals and Alloys* 19 (1944) 346
[37] ALLSOP, H. and STEVEN, W. *Iron and Steel Inst. Special Report No.* 36 (1946) 199
[38] JONES, F. W. and PUMPHREY, W. I. *J. Iron Steel Inst.* 157 (1947) 78
[39] SHEPHERD, B. F. *Trans. Amer. Soc. Metals* 22 (1934) 979
[40] ——*Trans. Amer. Soc. Steel Treating* 17 (1930) 90
[41] LIEDHOLM, C. A. *Metal Progress* (1944) 94
[42] ALLEN, N. P., PFEIL, L. B. and GRIFFITHS, W. T. *Iron and Steel Inst. Special Report No.* 24 (1939) 369
[43] COOPER, W. E. and ALLEN, N. P. *Iron and Steel Inst. Special Report No.* 36 (1946) 267
[44] BAIN, E. C. *Trans. Amer. Soc. Steel Treating* 20 (1932) 385
[45] SCHEIL, E. *Archiv. Eisenhuttenwesen* 8 (1935) 565
[46] See DELBART, G. *Metal Treatment* 14 (1947–8) 202
[47] GRANGE, R. A. and KIEFER, J. M. *Trans. Amer. Soc. Metals* 29 (1941) 85
[48] STEINBERG, S. *Metallurgy* 13 (1938) 7
[49] HOLLOMON, J. H., JAFFE, L. D. and NORTON, M. R. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 419
[50] JAFFE, L. D. *Metals Technology*, T.P. 2290, 1947
[51] RUSSELL, J. E. *Symposium on Internal Stress in Metals and Alloys, Institute of Metals* (1947) 95
[52] PUMPHREY, W. I. and JONES, F. W. *Iron and Steel Inst. Special Report No.* 36 (1946) 47
[53] —— *J. Iron Steel Inst.* 157 (1947) 27
[54] —— and JONES, F. W. *Ibid* 159 (1948) 37
[55] DELBART, G. and POTASZKIN, R. *J. Iron Steel Inst.* 157 (1947) 527
[56] HOLLOMON, J. H., JAFFE, L. D., MCCARTHY, D. E. and NORTON, M. R. *Trans. Amer. Soc. Metals* 38 (1947) 807
[57] HONEYMAN, A. J. K. *Iron and Steel Inst. Special Report No.* 36 (1946) 253
[58] *Wrought Steels (Carbon and Alloy Steels)* B.S. 971, British Standards Institution, 1944
[59] DEARDEN, J. and O'NEILL, H. *Trans. Inst. Welding* 3 (1940) 203
[60] BENNEK, H. and BANDEL, G. *Stahl und Eisen* 63 (1943) 653, 673, 695
[61] *U.K. Patent No.* 545,102, 30 July 1942
[62] STEVEN, W. *J. Iron Steel Inst.* 149 (1944) 239
[63] DONOHO, C. K. and MCCULLOCH, W. W. *Metal Progress* 44 (1943) 440 D
[64] Bethlehem Hardenability Calculator, Bethlehem Steel Company, Pa.
[65] BARDGETT, W. E. *Iron and Steel Inst. Special Report No.* 36 (1946) 81
[66] ALLSOP, H. *Ibid* 56

[67] CRAFTS, W. and LAMONT, J. L. *Trans. Amer. Inst. min. met. Engrs* 158 (1944) 157
[68] KRAMER, I. R., HAFNER, R. H. and TOLEMAN, S. L. *Ibid* 158 (1944) 138
[69] —— SIEGEL, S. and BROOKS, J. G. *Ibid* 167 (1946) 670
[70] BROPHY, G. R. and MILLER, A. J. *Ibid* 167 (1946) 654
[71] CRAFTS, W. and LAMONT, J. L. *Ibid* 167 (1946) 698
[72] —— —— *Ibid* 172 (1947) 303
[73] COMSTOCK, G. F. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 719
[74] GRANGE, R. A. and GARVEY, T. M. *Trans. Amer. Soc. Metals* 37 (1946) 136
[75] GLEN, J. *Iron and Steel Inst. Special Report No.* 36 (1946) 356
[76] HOLLOMON, J. H. and JAFFE, L. D. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 601
[77] HOUDREMONT, E. and SCHRADER, H. *Archiv. Eisenhuttenwesen* 5 (1932) 523
[78] SMOLUCHOWSKI, R. *Phys. Rev.* 62 (1942) 539
[79] HAWKES, M. F. and MEHL, R. F. *Amer. Inst. min. met. Engrs* 172 (1947) 467
[80] BUCKNALL, E. H. Discussion to reference 70
[81] LYMAN, T. and TROIANO, A. R. *Trans. Amer. Soc. Metals* 37 (1946) 402
[82] MANNING, G. K. and LORIG, C. H. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 442
[83] The Physical Properties of a Series of Steels, Part II. *J. Iron Steel Inst.* 154 (1946) 83
[84] MEHL, R. F. Hatfield Memorial Lecture 1948. *J. Iron Steel Inst.* 159 (1948) 113
[85] LORIA, E. A. and SHEPHARD, H. D. *Trans. Amer. Soc. Metals* 40 (1948) 758
[86] ALLEN, N. P. *J. Iron Steel Inst.* 157 (1947) 62
[87] HODGE, J. M. and OREHOSKI, M. A. *Trans. Amer. Inst. min. met. Engrs* 167 (1946) 502, 627
[88] JANITSKY, E. J. and BAEYERTZ, M. *Amer. Soc. Metals Handbook* 1939, p 515
[89] PATTON, W. G. *Metal Progress* 43 (1943) 726
[90] Section III *Iron and Steel Inst. Special Report No.* 36, p 23
[91] DIGGES, T. G. *J. Research Nat. Bur. Standards* 20 (1938) 571; *Trans. Amer. Soc. Metals* 28 (1940) 575; 29 (1941) 285; 30 (1942) 1
[92] GRENINGER, A. B. *Trans. Amer. Soc. Metals* 30 (1942) 1
[93] MCQUAID, H. W. and MCMULLAN, O. W. *Trans. Amer. Soc. Steel Treating* 16 (1929) 860
[94] SYKES, W. P. and JEFFRIES, Z. *Ibid* 12 (1927) 871
[95] GROSSMANN, M. A. *Principles of Heat Treatment* New York, 1935

BIBLIOGRAPHY

GROSSMANN, M. A., ASIMOW, M. and URBAN, S. F. *The Hardenability of Alloy Steels* New York, 1939
FRENCH, H. J. and LAQUE, F. L. *Alloy Constructional Steels* New York, 1942
BAIN, E. C. *Functions of the Alloying Elements in Steel* New York, 1945
Iron and Steel Inst. Special Report No. 36 1946
HOLLOMON, J. H. and JAFFE, L. D. *Ferrous Metallurgical Design* New York, 1947
MEHL, R. F. Hatfield Memorial Lecture 1948. *J. Iron Steel Inst.* 159 (1948) 113

6

PREFERRED ORIENTATION IN NON-FERROUS METALS

T. Ll. Richards

THE FOLLOWING is a general account of preferred orientation of crystals in metals, with special attention to the influence of cold rolling and annealing on the structure and properties of copper and brass strip.

Preferred orientation of crystals in metals may arise from several different causes. In castings, maximum crystal growth occurs in and along a particular crystallographic axis in the direction of maximum temperature gradient. Metal deposited electrolytically or by vacuum distillation takes up certain definite orientations. Hot and cold working also develops a pronounced preferred orientation or texture which may or may not be eliminated by subsequent annealing. The presence of well defined textures leads to directional properties in the final product whether it be in the form of extruded sections, wire, or strip. In extruded sections and wire it is possible that the textures might result in enhanced mechanical properties in the longitudinal direction but in non-ferrous metal strip complete absence of preferred orientation and consequent directionality are desirable.

Since an enormous quantity of strip is processed for subsequent fabrication by deep drawing, the study of preferred orientation in strip is of considerable practical importance.

In recent years there has been a development of x-ray diffraction and other techniques for the study of preferred orientation in worked and annealed metal. Although the primary object of these studies has been to determine the factors which give rise to directionality so that it may be effectively controlled, much information has also been obtained on the behaviour of metal during plastic deformation and subsequent recrystallization.

METHODS OF DETERMINING PREFERRED ORIENTATION

x-Ray and optical methods of determining the preferred orientation of crystals in metal specimens are fully described in standard works,[1,2] and so is the stereographic projection or pole figure. The pole figure

is essential if a full representation of the crystalline orientation is to be obtained. In the x-ray transmission method of establishing the pole figure a specimen is prepared of suitable thickness by dissolving in acid. A convenient thickness would be about 0·002 to 0·004 in (0·0508 to 0·1060 mm) for copper and copper base alloys and from 0·010 to 0·050 in (0·254 to 1·270 mm) for aluminium and its alloys. A series of transmission diffraction patterns are recorded on a flat film with the sample, for example a rolled strip, inclined at various angles to an incident x-ray beam, starting with the incident beam normal to the specimen and turning it 10° at a time, first about the rolling direction as axis and then about the transverse direction. The full radiation from a molybdenum or silver target is generally used since the characteristic wavelengths are short and a number of the low order rings diffracted by the more important crystallographic planes, whose orientations are to be plotted on the pole figure, can be recorded simultaneously on the photographic film.

The method outlined is satisfactory for most general purposes when the sample has a fine grain structure, but if the grain size is large it will be necessary to scan the specimen during exposure by a translation in its own plane with or without a small oscillation about the rotation axis. With really large crystals, greater than about 0·2 mm, it would probably be better to use an optical method.[3] By using flat specimens for transmission photographs the thickness of metal traversed by the x-ray beam increases with the angular deviation of the normal of the specimen from the beam direction and allowance must be made for absorption. BAKARIAN[4] has overcome this difficulty by machining cylindrical specimens so that the path of the beam in the specimen is independent of its angular position. Another weakness of the x-ray transmission method described is that the pole figure is discontinuous, since observations are made at 10° intervals. A continuous record, however, may be obtained by interposing a shield between the specimen and film which admits only the diffraction ring required and moving the film in a direction normal to the incident beam and specimen rotation axis at a rate proportional to the rate of rotation of the specimen. By this means it is theoretically possible to obtain a complete and continuous pole figure from a single photograph. With the transmission technique, however, some difficulty will be experienced in recording holes lying close to the normal specimen.

Wooster[5] has developed a method for plotting a complete pole figure from a single photograph by means of a camera in which the sample is set with a plane surface at the appropriate Bragg angle for the particular crystal planes to be recorded, and a cylindrical screen with a circumferential aperture admitting the cone of reflected x-rays placed before a moving cylindrical film mounted coaxial with the screen and incident x-ray beam. The sample is rotated in its own plane, its motion being geared to that of the film which is parallel to its axis. In this way the poles around the normal of the specimen are most easily recorded.

OCCURRENCE OF PREFERRED ORIENTATION

Preferred Orientation of Crystals in Castings and Deposited Films—Northcott and Thomas[6] prepared long columnar crystals in ingots of a number of copper rich alloys by cooling the base of the ingot only. The crystals were shown to be oriented with a cubic or ⟨100⟩ axis parallel to the major axis of the crystal. By suitable etching the dendritic or cored structure was revealed as striations in the polished section which were shown to be coincident with traces of {100} planes. Since three sets of such planes exist for each crystal three sets of striations could sometimes be observed, although those parallel to the long axis of the crystals were generally the most marked. It would seem that columnar crystals are formed by dendritic growth in the direction of the maximum temperature gradient, and that the main stem of the dendrite was parallel to a ⟨100⟩ axis with side arms also in ⟨100⟩ directions.

In electrodeposits the first deposited layer generally has a microstructure similar to that of the basis metal provided that the surface is clean; the crystal structure is of the same type and the lattice parameters are not widely different. In thick deposits, however, a preferred orientation is generally observed in which the crystals are oriented with an important crystal axis normal to the plated surface but randomly oriented about the normal. Thus, with deposits of metals having a face-centred cubic structure, orientations have been observed in which the crystals have either one or more of the following axes, namely ⟨100⟩, ⟨110⟩, ⟨111⟩, and ⟨112⟩, in the direction of the normal. Structures in which crystals are oriented with a crystal axis in a particular direction and are randomly oriented about this direction are termed fibre textures.

Sputtered and evaporated layers have similar fibre textures to those of electrodeposits, and textures of the ⟨100⟩, ⟨110⟩, and ⟨111⟩ type have been observed for face-centred cubic metals.

Preferred Orientation after Plastic Deformation—Crystal orientation developed by plastic deformation is of more general interest since this process is such a common feature of metal fabrication. Preferred orientation is developed by both hot and cold working, and the type for a particular metal is more dependent upon the geometrical change in shape than upon the manner in which the deformation is effected. For instance, the preferred orientation developed in a polycrystalline copper wire is the same when the wire is produced by similar reductions in cross-section by drawing through a die or by swaging. A preferred orientation of the same type as in drawn wire is produced even in a polycrystalline wire when it is extended in a tensile test.

Cold working results in work hardening of the metal as well as in a development of preferred orientation. Hot working, on the other hand, does not cause work hardening, but since the deformation is often effected at temperatures above that of recrystallization the ultimate texture corresponds to a recrystallization texture. Aluminium alloys, however, are generally hot rolled, forged, or extruded at temperatures below the recrystallization temperature so that a true deformation texture is developed. For example, when the duralumin type alloy is extruded as a cylindrical rod, the same type of preferred orientation is developed as in cold drawn rods of face-centred cubic metals.

TEXTURES AFTER PROCESSING

Textures in Drawn Wire—In wire drawing the metal crystals are deformed more or less in direct proportion to the deformation of the wire as a whole and the initially equiaxial crystals are elongated into long filaments or fibres parallel to the wire axis. The crystals also take up a definite orientation so that a particular crystal axis is parallel to the wire axis but quite randomly oriented about the axis, forming a very well defined fibre texture. Two types of orientation have been observed in varying proportions in drawn wires of face-centred cubic metals, namely with a ⟨100⟩ or ⟨111⟩ axis parallel to the wire axis. SCHMID and WASSERMANN[7] quote relative proportions for different metals as in the following table:

Table I. Proportions of ⟨100⟩ and ⟨111⟩ textures in Drawn Wires

Metals	Percentage of Crystals	
	With ⟨100⟩ parallel to Wire Axis	*With ⟨111⟩ parallel to Wire Axis*
Aluminium	0	100
Copper	40	60
Gold	50	50
Silver	75	25

It is not clear why the textures developed in drawn wire of metals all of which have the face-centred cubic structure should be so different. The ⟨111⟩ texture, however, appears to be dominant, and in fact with extremely large reductions in cross-sectional area there is evidence that the texture of copper is exclusively of the ⟨111⟩ type.

The wire or tension texture of polycrystalline samples appears to bear no relation to the stable end position of a face-centred cubic metal single crystal deformed in tension by duplex slip, namely a ⟨112⟩ axis parallel to the tension axis. It is significant, however, that for both the ⟨111⟩ and ⟨100⟩ textures there is a symmetrical distribution of the slip planes with respect to the wire axis, the former being apparently more favourably disposed to the direction of flow.

Textures in Cold Rolled Strip—The preferred orientation of crystals developed during the cold rolling of strip can best be represented by means of a pole figure, on which is plotted the distribution of poles of one of the principal crystallographic planes with respect to the rolling plane as the basic circle and the rolling direction as rotation axis. *Figures 1* and *2* are the ⟨111⟩ pole diagrams of heavily cold rolled copper and brass respectively. It is obvious from these diagrams that there is appreciable spread of the crystal orientation about more than one preferred position. Thus the crystals in heavily cold rolled brass (*Figure 2*) are spread about two preferred or ideal orientations, for which the common (110) plane is parallel to the plane of the strip with the $[\bar{1}12]$ axis parallel to the rolling direction for one orientation and the $[1\bar{1}2]$ axis for the other. The ideal positions of the ⟨111⟩ poles for the two orientations are indicated by letters with suffixes 1 and 2 in *Figure 2*. The poles suffix 1 are mirror images of the poles with

suffix 2, about planes normal to the strip surface which are either parallel or transverse to the rolling direction. Since the latter are {111} planes the two ideal orientations bear a twin relationship. Thus cold rolled brass strip is said to have a twin (110) [112] texture. From the shape of the pole regions of the twin textures the maximum spread from the ideal positions appears to be equivalent to a rotation of

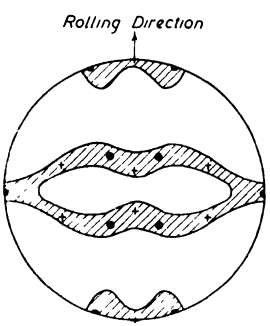


Figure 1. ⟨111⟩ *Pole figure of copper strip cold rolled to 97 per cent reduction. The dots represent* ⟨111⟩ *poles of twin* (110) [112] *textures, and the crosses* ⟨111⟩ *poles of twin* (112) [111] *textures*

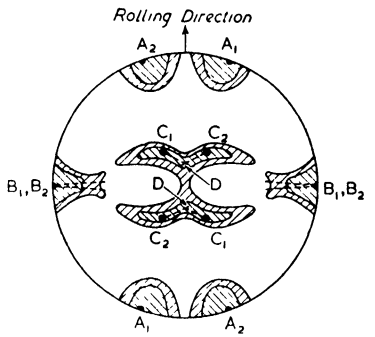


Figure 2. ⟨111⟩ *Pole figure of 70/30 brass cold rolled to 95 per cent reduction. The dots represent* ⟨111⟩ *poles of twin* (110) [112] *textures*

approximately 25° about their respective ⟨111⟩ poles A, which are closest to the rolling direction. The displacement of ⟨111⟩ poles by such rotations, which are indicated by broken lines in *Figure 2*, accounts for the merging of the ⟨111⟩ pole regions near the centre of the diagram and the splitting of the ⟨111⟩ pole regions B at the ends of the diameter transverse to the rolling direction.

The pole figure of the rolling texture of copper, *Figure 1*, not only consists of a spread of the twin (110) [112] textures but also of twin (112) [111] orientations. The latter twin texture is present to a lesser degree in rolled strip than the former, but pole figures have not yet been established with sufficient attention to intensity of x-ray reflections for different angular settings of the specimen for any precise estimate to be made of the relative proportions of the two types of twin textures. It seems likely that this estimate will be forthcoming from pole figures determined from a single pattern using the moving film technique.

The central ⟨111⟩ pole region in the rolled texture of 70/30 brass, *Figure 2*, is explained in terms of a rotation or spread of the C poles of

twin (110) [112] textures about their respective A poles as axes. By such rotations the C poles meet at points D which are almost coincident with the central ⟨111⟩ poles of the twin (112) [111] textures, suggesting that the latter is developed by a spread and merging of the main (110) [112] twin textures.

The rolling texture of most face-centred cubic metals is quoted as mainly (110) [112] plus (112) [111]. Other textures such as (100) [001] are sometimes quoted but it is likely that this texture is a residue of a recrystallization texture existing at an earlier stage in the process.

THEORY OF ROLLING TEXTURE

Theories of the development of rolling texture of face-centred cubic metals are not entirely satisfactory. WEVER[8] regards rolling as equivalent to a compression normal to the strip surface and a tension in the rolling direction. BOAS and SCHMID[9] and TAYLOR[10] have indicated how complex is the plastic deformation of polycrystalline metals, and the rolling texture of face-centred cubic metals is not a simple combination of the tension and compression textures of polycrystalline specimens. It is significant, however, that the main rolling texture (110) [112] corresponds to a combination of the stable end positions attained by double slipping when a single crystal of a face-centred cubic metal is deformed by tension and by compression, namely a ⟨112⟩ axis parallel to the tension axis and a ⟨110⟩ axis parallel to the compression axis respectively.

This correspondence of the behaviour of polycrystalline strip to that of a single crystal lends support to the theory of rolling textures developed by PICKUS and MATHEWSON[11] in spite of criticism by ELAM[11] on the grounds that plastic deformation of a polycrystalline metal is more complex than that of a single crystal. Pickus and Mathewson suggest that certain orientations are stable because the operating slip directions are symmetrical with respect to the direction of flow. This condition is also satisfied by the textures developed in cold drawn wire of face-centred cubic metals. Six possible rolling textures were predicted by Pickus and Mathewson, namely (110) [112], (100) [001], (110) [001], (112) [111], (010) [101], and (101) [101] in that order of relative stability. It is significant that VACHER[12] observed all except the last of these as intermediate textures developed when single crystals of copper of different initial orientations were cold rolled.

With very heavy cold rolling, however, the combined (110) [112] and (112) [111] textures were ultimately developed. Whereas the (110) [112] texture is given as the one of the greatest stability, the (112) [111] texture is fourth.

Progressive Development of Preferred Orientation—The rolling textures already described are the stable end textures developed after very heavy rolling reductions, generally 95 per cent or more, which are far greater than the reductions normally used in commercial practice. The magnitude of the reduction may be limited by work hardening of the material, but for metals such as copper or cupro-nickel, which can be cold rolled almost indefinitely, reductions must be restricted within specified limits to prevent the development of an undesirable structure.

It has been shown[13] that when polycrystalline copper strip is cold rolled, the individual crystals of which it is composed are deformed in more or less direct relation to the reduction of the strip itself, so that equi-axed crystals of 0·03 mm mean diameter are reduced to a mean thickness of 0·015 mm by a rolling reduction of 50 per cent and elongated to a mean length of 0·06 mm in the rolling direction. In the initial stages at least the deformation proceeds by a slip mechanism.

In polycrystalline face-centred cubic metal, as TAYLOR[10] has explained, five of the twelve available slip systems are operative if neighbouring grains are to remain contiguous throughout the deformation. The five slip systems will be distributed between three slip planes so that slip occurs in two directions on each of two planes and in one direction only on a third.

If the deformation of any one grain proceeds mainly by double slip operating on one plane, then there will be a rotation of the most active slip plane towards the plane of rolling about an axis transverse to the rolling direction. The rotation will continue with increased rolling until a second slip plane becomes more favourably oriented with respect to the deforming stresses. VACHER[12] has actually observed such rotation in the initial stages of the cold rolling of single crystal specimens of copper. There is, however, little evidence from x-ray or metallographic observations that slip deformation has occurred within individual crystals during the cold rolling of a strip of a pure metal. Asterism in x-ray transmission photographs of cold rolled copper strip may be interpreted as lattice disorientation due to slip

deformation within individual crystals according to its own particular slip systems. Slip lines and traces of slip planes become visible on the polished surfaces of polycrystalline metal specimens when they are plastically deformed. They are not revealed in specimens of cold rolled copper[13] which are sectioned, polished, and etched after deformation but can be observed in specimens of cold rolled brass[14] similarly prepared.

There appears to be a limit to the amount of deformation produced by slip, for after reductions in thickness in cold rolling of about 33 per cent for copper and slightly more for brass there is evidence that a second mechanism is operative. Deformation bands can be seen in sections of both copper and brass polished and etched after rolling with reductions in excess of the amounts indicated, from which it appears that deformation in this stage has occurred by a break-up of crystals on surfaces lying close to planes of maximum resolved shear with respect to the rolling loads. These surfaces make traces in the rolling plane which are transverse to the rolling direction and are inclined at an angle somewhat less than 45° to the surface. The character of the deformation layers has never been properly elucidated. They are rather irregular and it is consequently difficult to associate them with any specific crystallographic plane. It is significant, however, that they lie parallel to certain ⟨110⟩ directions of the (110) [112] and (112) [111] textures, and it is possible that the deformation bands resemble those in iron for which the deformation proceeds in the most favourable slip direction without reference to any slip plane.

A certain amount of preferred orientation is developed by deformation of the first type, but since each crystal within a polycrystalline mass is deforming largely according to its own particular slip system the degree of preferred orientation developed is rather small. Marked preferred orientation is developed during the second stage of deformation and the degree of preferred orientation increases progressively with increased amounts of cold rolling but is independent of the grain size of the slip before rolling.

ANISOTROPY AFTER PLASTIC DEFORMATION

The development of preferred orientation in cold rolled metals produces some degree of mechanical anisotropy, but in general the magnitude of the resulting variation in properties is not comparable with the change

brought about through work hardening and by the presence of the deformation layers described in the previous section. Thus the tensile strength of copper[13] is increased from 15 tons/sq in (23·6 kg/mm²) in the soft condition to about 30 tons/sq in (47·2 kg/mm²) after a cold rolling reduction of 97 per cent, but the anisotropy amounts only to a diminution of tensile strength by 2 tons/sq in (3·1 kg/mm²) in the case of test pieces taken at 45° to the rolling direction as compared with the values for test pieces at 0° and 90°, the latter values being identical. It is noteworthy that slight differences between the tensile strength values at 0° and 90° to the rolling direction in cold rolled copper are observable after reductions ranging from 25 per cent to 90 per cent, the latter being slightly greater and the minimum value still occurring at 45°.

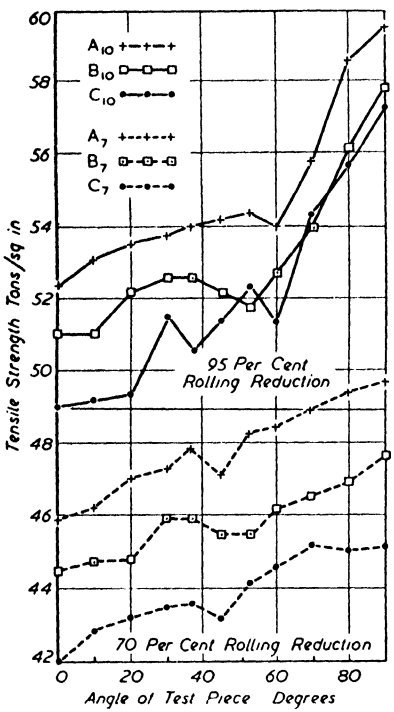


Figure 3. The variation of tensile strength with direction of testing for 70/30 brass strip of three initial grain sizes

The anisotropy of cold rolled brass[14] strip is somewhat different from that observed for copper strip, contrary to expectations in view of the similarity of the rolling texture of these metals. For brass strip the tensile strength is least in the direction of rolling and greatest in the transverse direction, the difference in values increasing with the amount of cold rolling. The variation in tensile strength with direction of testing in 70/30 brass strip of three initial grain sizes, cold rolled with reductions in thickness of 70 and 95 per cent, is shown in *Figure 3*.

The difference in tensile strengths in the directions at 0° and 90° to that of rolling can be explained, at least qualitatively, for both copper and brass by the presence of deformation layers, although these appear to have a greater influence in the case of brass. With rolling reductions of 50 per cent and more deformation proceeds by crystal break-up on the deformation layers or planes situated transverse to the rolling direction and inclined a little less than 45° to the strip surface. In

subsequent rolling or in tensile testing in the rolling direction deformation occurs by shear on the deformation layers. The resolved shear on these layers due to an applied tensile stress is a maximum for tension in the rolling direction and a minimum for tension in the transverse direction. This effect is common to many heavily rolled metals and alloys, and in some instances excessive cold rolling leads to a complete disruption of the material by fracture on the deformation layers, as illustrated by the photograph of a piece of over-rolled aluminium alloy reproduced in *Plate XXXV*.

Single crystals of face-centred cubic metals are anisotropic in respect of their elastic properties, Young's modulus having a maximum value in a ⟨111⟩ direction and a minimum in a ⟨100⟩ direction. The degree of anisotropy, according to SCHMID and BOAS,[15] is quite considerable in copper, silver, and gold but only slight in aluminium.

The elastic modulus of a metal, however, is not greatly changed by plastic deformation, although the elastic limit of annealed material is extended by cold working. Observation of the elastic modulus then provides a possible way of examining the effect of preferred orientation independently of any work hardening effect, provided that no mechanical defects are present. WEERTS[16] has calculated the Young's modulus values of heavily rolled copper strip in various directions, assuming the rolled structure to be composed of equal amounts of the (110) [112] and (112) [111] textures. The values so obtained were in reasonable agreement with those determined experimentally. The Young's modulus was a maximum in the directions at 0° and 90° to the rolling direction and a minimum in the 45° direction. The modulus of cold rolled copper thus varies in a similar manner to its tensile strength.

Since the structure of drawn copper wire is generally composed of two textures for which the directions of maximum and minimum modulus respectively are parallel to the wire axis, then cold drawing should not be expected to affect the modulus value of wire to any appreciable extent. For very large reductions in cross-sectional area, however, there is evidence that the proportion of the ⟨100⟩ fibre texture diminishes with a corresponding increase of the ⟨111⟩ texture, which should result in a progressive increase in the modulus values. This may account in part for the high tensile strength values observed for the central zones in drawn copper wire in which the preferred orientation is most highly developed.[7]

The hot working or extrusion of the duralumin type alloy[17] results in a high degree of preferred orientation but is not accompanied by any work hardening effects. The anisotropy in mechanical properties of the material is then directly related to the preferred crystalline orientation. The tensile strength is greater in a direction parallel to the axis of an extruded rod than in the transverse direction, probably due to the development of a predominant ⟨111⟩ texture in the aluminium alloy rod during extrusion. The texture of the rod is unaltered by subsequent solution heat treatment and room temperature age hardening apart from the development of a skin of coarse grain structure at the surface.

The mechanical properties of the material are enhanced by the heat treatment, while the tensile strength remains greater parallel to the rod axis, the value in this direction being in fact appreciably greater than in rod with a random orientation of crystals.

In the manufacture of forgings such as airscrews from extruded sections of light alloy, where maximum tensile properties are required, the hot working conditions are controlled so that the fibre structure is maintained after final solution heat treatment. Forging at too low a temperature results in some degree of cold working, with consequent lowering of the recrystallization temperature below that of solution heat treatment.

RECRYSTALLIZATION TEXTURES

Recrystallization textures are generally the result of annealing material with a pronounced deformation texture. They are not usually as sharply defined as deformation textures because the grain size of annealed material is appreciable and there is some scatter of the orientation of individual crystals about the mean position. In the x-ray examination of annealed material separate diffraction spots will be obtained from each diffracting crystal, and if the grain size is greater than about 0·02 mm mean diameter there will be, with stationary specimens, insufficient diffraction spots to form a recognizable pattern. Some improvement may be obtained by translating the sample in a fixed plane during exposure so that a greater number of crystals can contribute to the pattern.

A knowledge and control of recrystallization textures is of considerable importance to metal strip manufacturers, particularly if the material

is to be used for press work when directionality is to be avoided. The study of preferred orientation in annealed material is also of fundamental as well as practical interest, for recrystallization textures are, more often than not, quite different from the deformation textures from which they are derived. A comparison of the pole figures of textures of rolled and annealed material should give some insight into the mechanism of the recrystallization process. The state of the theory of the formation of recrystallization textures is less satisfactory than that for deformation textures, and much further work is necessary in this field.

Texture in Annealed Wire—BARRETT[1] reports that there is no change from the deformation texture when aluminium is annealed at temperatures below 500°C nor when copper is annealed at temperatures below 1000°C. The texture of aluminium and copper wire annealed at higher temperatures is uncertain owing to coarse grain growth.

Textures in Annealed Strip—A texture similar to that in the cold rolled strip has been reported by WILSON and BRICK[18] for 70/30 brass cold rolled with a reduction in thickness of 97 per cent and finally annealed for one hour at 650°C. Most recrystallization textures in annealed strip, however, differ from those developed by cold rolling, and the most common texture reported for face-centred cubic metals is (113) [211]. This was first observed by GLOCKER[19] in rolled and annealed silver, and later by GOLER and SACHS[20] in some copper, silver, and gold alloys. The (113) [211] texture in rolled and annealed brass has been confirmed many times, most recently by BRICK,[21] COOK and RICHARDS[14] and WILSON and BRICK.[18] It is probably a true recrystallization texture since the annealing temperature adopted in these latter experiments was as low as 400°C in order to have specimens of small grain size. Wilson and Brick[18] observed the same texture in brass cold rolled with a reduction of 98 per cent and subsequently annealed for 24 hours at 250°C. These authors attribute the change back to the (110) [112] deformation texture on annealing at 700°C to grain growth.

Another characteristic recrystallization texture in annealed strip of face-centred cubic metals is (100) [001]. A list is given by BARRETT[1] of the metals for which this texture has been observed. These include aluminium, copper, gold, nickel, and the iron-nickel alloys which have a face-centred cubic structure, that is those containing 30 to 100 per cent nickel.

The effect of various solute elements on the recrystallization texture of copper have been studied in some detail by BRICK, MARTIN, and ANGIER.[22] These workers note that an addition of only 1·2 atomic per cent zinc was sufficient to change the texture from that of copper to approximately that of brass. Additions of aluminium, cobalt, manganese, nickel, nickel-zinc, and iron had the similar effect of changing the recrystallization texture from that of copper to that of brass. It would seem, however, from the pole figure for the alloy with 5·48 per cent nickel that there was an appreciable amount of the (100) [001] texture in the annealed material. MÜLLER[23] has studied recrystallization textures for alloys in the ternary system iron-nickel-copper cold rolled with a reduction of 99 per cent. An extensive range of nickel rich alloys recrystallized with the (100) [001] texture, including all the binary copper-nickel alloys. The difference in the observations of Müller and Wilson and Brick is probably due to the differences in the processing conditions upon which the type of directionality developed in the finished strip is dependent, as KAISER[24] showed for copper and BASSETT and BRADLEY[25] for cupro-nickel.

Development of Recrystallization Textures—The two main textures described in the preceding section are generally the consequence of very heavy rolling reductions and subsequent annealing, but strip manufacturers require to know what specific variations in processing conditions are responsible for the development of recrystallization textures, in view of their practical importance.

BALDWIN,[26] and COOK and RICHARDS[13] have studied in detail the effect of variations in initial structure, final rolling reduction, and final annealing temperature on the structure of copper strip, while the latter investigators considered also the influence of variation in the grain size of the material before the final rolling reduction. Both investigations were made on tough-pitch or oxygen bearing coppers of very similar quality, although Baldwin examined both electrolytic tough-pitch copper and silver bearing tough-pitch copper and found that the structural changes associated with variations in processing conditions were practically identical for the two types of copper. Considering that these investigations were carried out quite independently during the war years, the agreement between the observations is most satisfactory.

Cook and Richards prepared four series of strips each representing a distinct type of initial structure and three sets in each series with a

particular average grain size. These strips were then cold rolled with various reductions over a wide range and finally annealed at a number of different temperatures. It was observed that the type of structure present at the initial stage had little effect on the behaviour of the copper on subsequent rolling and annealing, so that observations on three series of strips were very much the same and only the observations on series N in which the initial crystals were randomly orientated need be described as typical for the three series. Strip of the fourth series S in which the initial structure was the recrystallization texture (100) [001] behaved somewhat differently. Differences in initial grain size were found to have a more significant effect on the final structure than differences in initial structure.

With the three sets of strips of the N series it was found that cold rolling with reductions in excess of 50 per cent developed a preferred orientation corresponding to the usual rolling texture of copper, now specified as (110) [112] and (112) [111]. The degree of preferred orientation was independent of the initial grain size and dependent only on the magnitude of the final rolling reduction. The individual crystals of set NA, NB, and NC of initial average grain size A 0·025 to 0·03 mm, B 0·04 mm, and C 0·06 to 0·065 mm were deformed during rolling in direct proportion to the reduction of the strip itself, so that the average thicknesses of the crystals were, after a rolling reduction of 90 per cent, 0·0025 to 0·003 mm, 0·004 mm, and 0·006 to 0·0065 mm respectively. x-Ray transmission photographs of the complete N series of strips after annealing for one hour at 400°C are reproduced in *Plate XXXVI*. These were taken using molybdenum radiation with the rolling direction vertical and the specimens set so that the incident beam was inclined at 10° to the normal to the strip about the rolling direction. By this means the presence of crystals with orientations parallel to the (100) [001] texture were readily detected, since one family of {100} planes would be in strongly diffracting positions.

The transmission photographs show that the grain size in the finally annealed strip decreases with increasing cold rolling reduction. On annealing after reductions less than 50 per cent there is a completely random orientation of crystals, but with reductions in excess of 50 per cent, however, there is a development of preferred orientation. For the set of strips NA the strong radial streak in the diffraction patterns *Plate XXXVI*, characteristic of (010) reflections of the (100) [001]

texture, first appears in the pattern of strip NA5 cold rolled with final reduction of 80 per cent and is more marked in the pattern of NA6 (reduction 90 per cent) and more or less fully developed in the pattern of NA7 (reduction 95 per cent). The complete development of the (100) [001] texture in the NA set of copper strips thus occurs on annealing at 400°C after a reduction of 95 per cent when the initial grains of average diameter 0·03 to 0·035 mm have been cold rolled to a mean thickness of 0·0016 mm. For the NB set of strips a reduction greater than 95 per cent but less than 97·2 per cent is necessary before the texture is completely of the (100) [001] type on annealing at 400°C, while the reduction required in the NC series is greater than 97·2 per cent. The main grain thickness corresponding to the critical reduction is practically the same for all three sets of strips. More precise determination[27] of this critical dimension for different temperatures of annealing showed that it varied linearly with temperature, having a value of about 25×10^{-5} cm at 900°C and extrapolating to a zero value at the temperature of absolute zero. The value of the critical grain thickness for development of the (100) [001] recrystallization texture at room temperature is 6×10^{-5} cm, and it has been shown[28] that when copper strip has been cold rolled so that the crystals are thinner than this amount then spontaneous annealing occurs at ordinary room temperature.

With strips of sets NB and NC which have been cold rolled with reductions in excess of 50 per cent but less than the critical reduction the preferred orientation developed on subsequent annealing is not very well defined but bears some resemblance to the texture in the rolled strip. The conclusion is then drawn that when the crystals have been cold rolled so that they are thicker than a certain critical dimension, then recrystallization occurs independently within each crystal. When, however, the crystals are thinner than the critical dimension after cold rolling, then recrystallization within any one crystal has an influence on neighbouring crystals and a merging of their textures occurs to form the (100) [001] recrystallization texture. The change in orientation from that of the copper rolling texture (110) [112], (112) [111], to the (100) [001] recrystallization texture cannot be considered as simple rotation or twinning but rather as a complex process of atomic rearrangement.

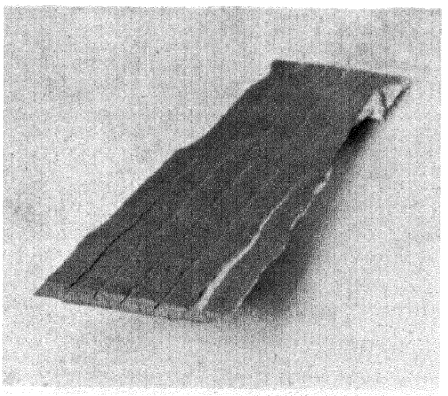


Plate XXXV. Over rolled aluminium alloy strip

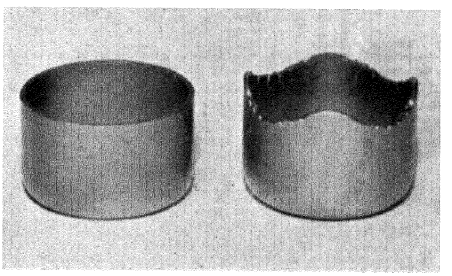


Plate XXXVII. Cylindrical cups made from isotropic and anisotropic copper strip, the latter with a (100) [001] recrystallization texture

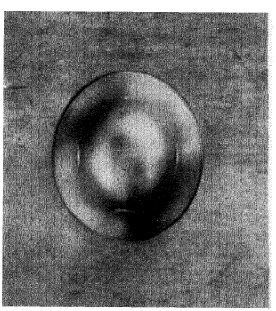


Plate XXXVIII. Erichsen test dome in anisotropic copper strip with (100) [001] recrystallization texture

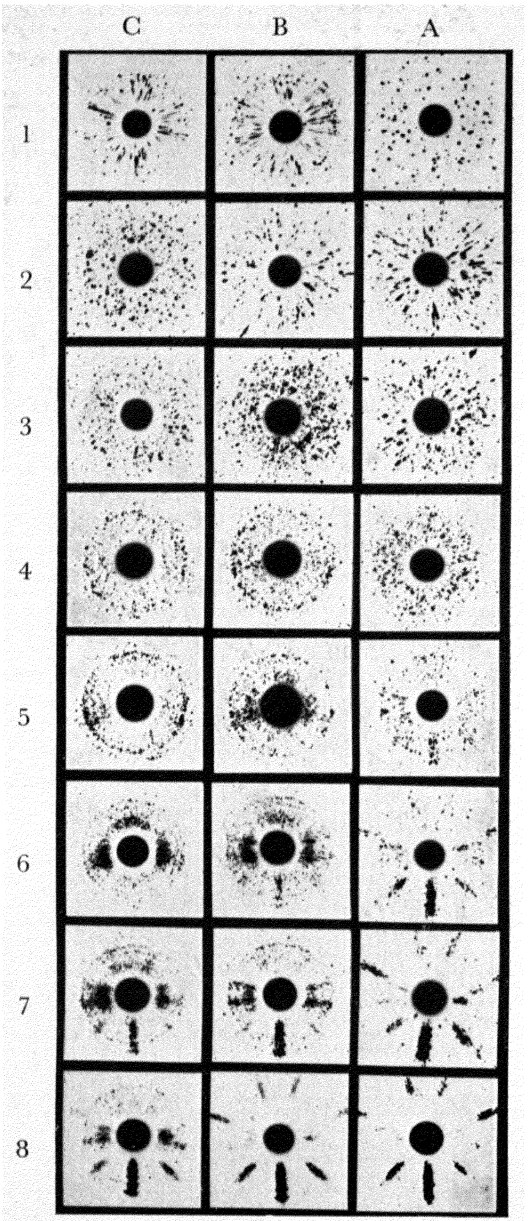


Plate XXXVI. X-ray transmission photographs (molybdenum radiation) of copper strips of series **N, cold rolled** *with the reductions indicated and annealed for one hour at 400°C*

SACHS and SPRETNAK[29] have expressed the view that the development of the (100) [001] texture is due to the presence of small amounts of the texture in rolled metal acting as points of nucleation. This explanation is not very satisfactory since it does not take into consideration the clear dependence of the process on the grain thickness after rolling, and seems quite untenable in the light of the behaviour of the copper strip of series S with a specially prepared (100) [001] texture on subsequent cold rolling and annealing. As forecast by PICKUS and MATHEWSON,[11] this texture was found to be very stable on rolling. It was hardly altered by reductions of 50 per cent, and there was evidence in diffraction patterns of a substantial amount of the (100) [001] after reductions in thickness up to 95 per cent in addition to the normal rolling texture of copper. On subsequent annealing, however, heavily rolled strip of the S series tended to develop a recrystallization texture of the (113) [211] rather than the (100) [001] type, so it is apparent that the presence of (100) [001] texture in the rolled strip does not contribute to the formation of the same texture in the finally annealed material.

COOK and RICHARDS[14] have also studied in detail the development of the recrystallization texture in 70/30 brass strip. In this work three sets of strip A, B and C were prepared with a random orientation of crystals but with three different initial grain sizes, 0·01 to 0·015 mm, 0·025 mm, and 0·06 to 0·065 mm respectively. These were cold rolled with reductions in thickness from 10 to 95 per cent and annealed finally at temperatures ranging from 400 to 800°C. In order to finish up with a series of strips all at the same final thickness, material was prepared at different gauges before the final rolling, which meant that strips rolled with small final rolling reductions had already been subjected to a number of cold rolling reductions and intermediate annealing. To obtain the desired initial grain size, strips of the A series were annealed at 400°C, cold rolled 25 per cent and again annealed at 400°C, while strips of the B and C series were processed similarly but annealed at 500°C and 575°C respectively.

x-Ray examination was carried out on samples of strip in the cold rolled condition, and after annealing at 400°C. The same (110) [112] rolling texture was developed in all three sets of strips to a degree which was independent of the initial grain size and dependent only on the magnitude of the final rolling reduction. The ⟨111⟩ pole figure of the

brass strip A10 cold rolled with a reduction of 95 per cent, which is reproduced in *Figure 2*, has already been described.

In brass strip annealed after rolling with reductions of 10 to 50 per cent there was a preferred orientation similar in character to the rolling texture, the behaviour of brass being therefore unlike that of copper strip in which there was a completely random orientation after processing under the same conditions. The degree of preferred orientation developed actually decreased with the magnitude of the final rolling reduction in the range 10 to 40 per cent and with the initial grain size, being greatest in strip A1 of smallest initial grain size rolled with a reduction of 10 per cent, and least in C4 of largest initial grain size rolled with 40 per cent reduction. It appears that in brass strip the preferred orientation developed in preliminary stages is not completely removed on subsequent annealing, particularly if the reductions are small and the intermediate annealing temperature low.

With reductions in thickness in the range 40 per cent to 80 per cent there was in all three series of strips an increase in the degree of preferred orientation on annealing. The degree of preferred orientation developed was not marked and it was consequently difficult to determine the orientation precisely. It was greatest in strips of the A series of smallest initial grain size.

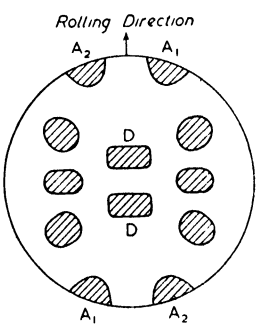


Figure 4. ⟨111⟩ *Pole figure of 70/30 brass cold rolled to 95 per cent reduction and annealed at 400°C*

BURGHOFF and BOHLEN[30] have published pole figures of the distribution of {111} planes in brass strips rolled with reductions of 50 and 84 per cent and subsequently annealed.

While there was a slight difference in the two pole figures, the preferred orientation represented by both approximated to the texture of the rolled strip.

Marked preferred orientation was observed only in the strips of the three series rolled with the two heaviest reductions in thickness, namely 90 and 95 per cent. The texture ultimately developed by heavy rolling and subsequent annealing which was identical in the strips of the three series, A, B and C, is represented by the ⟨111⟩ pole diagram reproduced in *Figure 4*. This is in complete agreement with that established by BRICK.[21]

By a direct comparison of the pole figures of cold rolled brass strip, *Figure 2*, and annealed strip, *Figure 4*, an analysis of the mechanism of reorientation on annealing can be made. The rolling texture is the twin (110) [112] and suffixes 1 and 2 in *Figure 2* relate to ⟨111⟩ pole regions corresponding to the separate twin textures. The A_1 and A_2 pole regions are common to both the rolling texture, *Figure 2*, and the recrystallization texture, *Figure 4*. The atomic rearrangement which occurs on annealing heavily cold rolled strip can then be considered as a rotation or twinning or both of the separate twin textures on the (111) planes corresponding to their respective A poles. A diagram illustrating such a mechanism is reproduced in *Figure 5*.

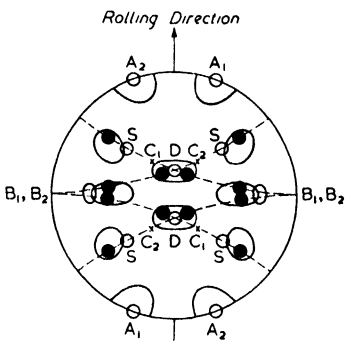


Figure 5. Diagram illustrating the atomic rearrangement which occurs on annealing heavily cold rolled strip. This can be considered as a rotation or twinning or both of the separate twin textures on the (111) planes corresponding to their respective A *poles*

By rotations of 22½° to the right and to the left the central pole regions C_1 and C_2 in the pole figure of the rolled strip merge to form the central pole regions D of the pole figure of the annealed strip and to form new pole regions in the position of the outlined circles marked S, *Figure 5*. Rotation on the (111) planes may give rise to twinning on those planes, and the positions of ⟨111⟩ poles of the rolling texture after 22½° rotations and subsequent twinning are indicated by the full black circles. These latter positions correspond more closely than those of poles resulting from rotations only to the observed pole regions for annealed strip, so that the reorientation on recrystallization of heavily rolled brass should be regarded as a combination of rotation and twinning of the twin (110) [112] textures of the rolled strip.

The mechanism of reorientation occurring on annealing heavily rolled copper, as stated earlier, cannot be considered as simple rotation and twinning as in the case of brass, but it is perhaps significant that the positions of poles marked S in *Figure 5* correspond to the positions of the ⟨111⟩ poles of the (100) [001] recrystallization texture.

ANISOTROPY OF ANNEALED MATERIAL

Anisotropy in annealed material is directly related to preferred orientation and is a factor of major importance in strip manufacture. Certain recrystallization textures are specially developed in silicon-iron or iron-nickel strip for transformer laminations so as to take advantage of the high magnetic permeability values obtaining in particular crystallographic directions. In non-ferrous metals, however, anisotropy or directionality is generally an undesirable feature which should be avoided, since it leads to difficulties in pressing or deep drawing operations.

The initial approach to the problem, notably by KAISER,[24] BASSETT and BRADLEY,[25] and COOK,[31] was largely on empirical lines. In these investigations metal strip was processed with rolling reductions of different magnitudes and with various intermediate annealing conditions. Anisotropy in the finally annealed strip was detected and estimated by a simple but sensitive cupping test, in which a circular blank is cut from the strip and formed into a cylindrical cup by passing through a die. If the strip is completely free from directional properties, then the cups so formed have level brims, but cups from anisotropic strip have wavy tops with the crests or ears situated symmetrically with respect to the rolling direction.

Kaiser[24] observed four ears situated at 0° and 90° or at 45° with respect to the rolling direction on cups cut from annealed copper strip, depending on the processing conditions. Bassett and Bradley[25] found that cupro-nickel strip had cupping characteristics similar to copper. The formation of ears of the 0°, 90° type has not been observed on cups made from brass strip, but the 45° type is quite common.[31] Earing has also been observed[14,18,30,32] in other positions on cups from brass strip, namely four ears at 55° and six ears at 0° and 60° to the rolling direction, depending on the rolling and annealing procedure. Kaiser,[24] Bassett and Bradley,[25] and Cook[31] all appreciated that the anisotropy of the strip was due to a preferred orientation of crystals. Kaiser was able to relate the 0° and 90° type of earing for copper strip to a preferred orientation of crystals such that a {100} plane of the face-centred cubic lattice of copper was parallel to the strip surface and a ⟨100⟩ axis parallel to the rolling direction, whereas the 45° type of earing was related to a preferred orientation of crystals in which a {110} plane was parallel to the strip surface. Bassett and

Bradley noted that cupro-nickel strip which yielded cups with 0° and 90° ears had a typical microstructure with twins aligned in one of two directions at 90° to each other. Cook observed the distribution of twinning directions in rolled and annealed brass strip and was able to connect directionality in properties with preferred crystalline orientation, although the preferred orientation was of a complex character. With strip having the (100) [001] recrystallization texture the ears on cups are invariably situated at 0° and 90° to the rolling direction. The height of the ears increases with the proportion of the crystals in the strip which are oriented similarly to the (100) [001] texture,[33] the ears being of appreciable height when the proportion exceeds about 10 per cent. Photographs of a flat topped and an eared cup from isotropic copper strip and anisotropic strip with the (100) [001] texture respectively are shown in *Plate XXXVII.*

A second effect arising from directionality is due to low ductility in certain crystallographic directions. This effect, which is of consequence only in strip with a high proportion of the (100) [001] texture, is illustrated in *Plate XXXVIII.* This is a photograph of the dome produced in an Erichsen test, in which a steel ball is pushed against a clamped copper strip until the first signs of fracture appear, the height of the dome being a measure of the ductility of the strip. With directional copper of (100) [001] texture, fractures appear prematurely in positions at 0° and 90° to the rolling direction.

The connection between the earing effect of anisotropic strip and the variation in its mechanical properties, tensile strength, and elongation is not obvious. Thus, for example, the crests of ears on cups from brass strip are situated at points on the brim corresponding to directions in the strip for which the tensile strength values are a minimum and elongation values are a maximum,[14] while from copper strip with the [100] [001] texture the ears are situated in the 0° and 90° positions corresponding to positions of maximum tensile strength and minimum elongation.[34]

This apparent lack of correlation of earing behaviour with mechanical properties is further illustrated by diagrams published by Wilson and Brick,[18] in which the values of various mechanical properties of copper and brass strips are plotted against the direction of testing and the variations compared with a record of the cup contour. The only really consistent correlation is that the elastic modulus values in tension

and compression are at a maximum for directions corresponding to the troughs of the wavy tops.

A rational interpretation of earing behaviour in terms of the structure has been made by Wilson and Brick, who compared the contour of cups from directional brass strip with the ⟨111⟩ pole figure of the strip. The troughs were in general observed in directions corresponding to a high concentration of ⟨111⟩ poles on the periphery of the pole diagram and ears in the direction of low peripheral concentrations. The connection between earing directions and peripheral poles is explained by assuming that in the cupping operation the major deforming stresses are circumferential, the effect of these being to reduce the diameter of the circular blank to that of the cylindrical cup. Plastic deformation in face-centred cubic metals occurs by slip on {111} planes, and circumferential stresses in directions of peripheral ⟨111⟩ poles develop no resolved shear on the corresponding {111} planes. The resolved shear is greatest in directions 45° from peripheral ⟨111⟩ poles, so that circumferential stresses should have maximum effect and should develop ears in these directions. To complete this explanation the behaviour of each component of the recrystallization texture should be considered independently and directions at 90° to peripheral concentrations of ⟨111⟩ poles as well as the directions of the poles themselves should by the same argument be positions of troughs on cups. The explanation does not take into account the influence of poles other than the peripheral ⟨111⟩ poles and so cannot predict the earing behaviour of copper strip with the (100) [001] recrystallization texture.

The (100) [001] recrystallization texture has only one component and the structure can be regarded as practically equivalent to that of a single crystal. A rigorous analysis of the problem of directionality in strip with the (100) [001] texture is then possible. The (113) [211] recrystallization texture of brass is far more complex, since it is made up of at least eight components, four arising from two opposite rotations of twin (110) [112] textures and four from subsequent twinning.

WEERTS[16] has shown that there is a very close agreement between experimentally determined Young's modulus values in different directions of a copper strip with the (100) [001] texture and theoretical values computed from the known maximum and minimum values in ⟨111⟩ and ⟨100⟩ directions respectively.

Several workers[26,34–37] have examined the directionality in

mechanical properties of copper strip with the (100) [001] texture and there is a reasonable measure of agreement in their results. Cook and Richards[34] have made a detailed analysis of their observations, in which the variations in mechanical properties with direction of testing are related to the resolved shear on the operating slip planes due to applied tensile forces. There are four families of {111} or slip planes, and the values of the shear on each of these resolved along the nearest slip direction due to an applied tensile stress are plotted in *Figure 6* against the direction of the applied stress. Also plotted in *Figure 6* are the experimentally determined elongation and tensile strength values.

Elongation is a minimum in directions at 0° and 90° to the rolling direction, that is in ⟨100⟩ directions, corresponding to directions of the applied stress for which the resolved shear on each of the four slip planes is high and equal in value. It is a maximum in the 45° or ⟨110⟩ directions, the direction of the applied stress for which the resolved shear is zero on two of the slip planes, and has a minimum value on the other two. The ductility of the strip is therefore least in the direction in which the applied stress causes all four sets of slip planes to operate simultaneously and greatest in the directions in which the applied stress results in slip on only two sets of planes. The directions of minimum ductility in tension thus correspond to positions of premature fracture in Erichsen tests, *Plate XXXVIII.*

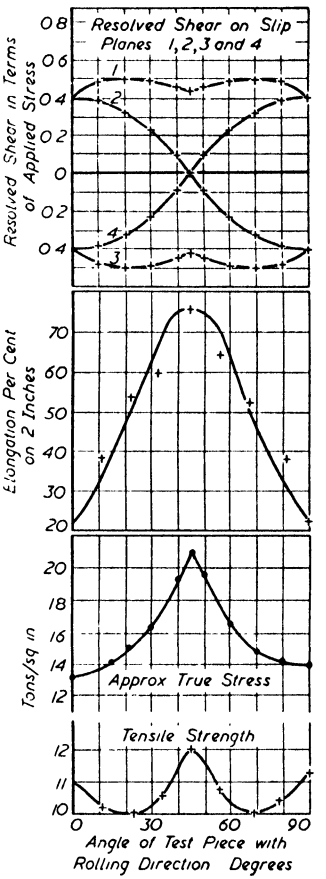


Figure 6. The variation of tensile strength, approximate true stress, elongation, and resolved shear with the angle of test piece with the rolling direction for copper strip

The tensile strength values appear to follow inversely the variations in the resolved shear on the planes of maximum resolved shear 1 and 3, or, in other words, it would seem a constant resolved shear is required

in the operating slip planes to cause fracture, for all directions of testing, which is analogous to the conditions for slip in single crystals.[10] This correspondence is more apparent than real, for tensile strength values are expressed in terms of maximum applied load divided by initial cross-sectional area of the test piece. The actual or true stress at fracture should be determined by dividing the maximum load by the reduced area of cross-section at point of fracture. An approximate estimation of the true stress at fracture can be made by multiplying the tensile strength value by a factor $(1 + e)$, where e is the elongation of the test piece expressed as a fraction of the gauge length. The approximate true stress values at fracture, also plotted in *Figure 6*, indicate that the variations in true stress values are similar to the variations in elongation values, being at a minimum in the 0° and 90° or ⟨100⟩ directions and at a maximum in the 45° or ⟨110⟩ directions. The true stress/strain characteristics of a metal in tension and compression are practically identical so that it would seem that the positions of ears on cups made from copper strip with the (100) [001] recrystallization texture correspond to directions of minimum true tensile or true compressive strength of the strip. The same relationship holds for annealed brass strip since the observed variation in elongation values are so small that the directions of maximum and minimum true stress values are identical with the positions of the maximum and minimum tensile strength values.

The writer wishes to express his thanks to Dr Maurice Cook for his encouragement and interest, and to his colleague Mr E. C. Williams for assistance in the preparation of the manuscript.

BIBLIOGRAPHY

SCHMID, E. and BOAS, W. *Kristallplastizitat* Berlin, 1935
BARRETT, C. S. *Structure of Metals* New York and London, 1943
TAYLOR, A. *An Introduction to X-ray Metallography* London, 1945

REFERENCES

[1] BARRETT, C. S. *Structure of Metals* New York and London, 1943
[2] TAYLOR, A. *An Introduction to X-ray Metallography* London, 1945
[3] —— *J. sci. Instrum.* 25 (1948) 301
[4] BAKARIAN, P. W. *Trans. Amer. Inst. min. met. Engrs* 147 (1942) 266
[5] WOOSTER, W. A. *J. sci. Instrum.* 25 (1948) 129
[6] NORTHCOTT, L. and THOMAS, D. E. *J. Inst. Met.* 65 (1939) 205
[7] SCHMID, E. and WASSERMANN, G. *Z. Physik* 42 (1927) 779
[8] WEVER, F. *Trans. Amer. Inst. min. met. Engrs* 93 (1931) 51

[9] BOAS, W. and SCHMID, E. *Z. tech. Phys.* 12 (1931) 71
[10] TAYLOR, G. I. *J. Inst. Met.* 62 (1938) 307
[11] PICKUS, M. R. and MATHEWSON, C. H. *Ibid* 64 (1939) 555
[12] VACHER, H. C. *J. Re. Nat. Bur. Stand.* 26 (1941) 385
[13] COOK, M. and RICHARDS, T. LL. *J. Inst. Met.* 67 (1941) 203
[14] —— —— *Ibid* 69 (1943) 351
[15] SCHMID, E. and BOAS, W. *Kristallplastizitat* Berlin, 1935
[16] WEERTS, J. *Z. Metallk.* 25 (1933) 101
[17] HARDY, H. K. *Metallurgia* 30 (1944) 240
[18] WILSON, F. H. and BRICK, R. M. *Trans. Amer. Inst. min. met. Engrs* 161 (1945) 173
[19] GLOCKER, R. *Z. Physik* 31 (1925) 386
[20] GOLER, F. v. and SACHS, G. *Ibid* 56 (1929) 485
[21] BRICK, R. M. *Trans. Amer. Inst. min. Engrs* 137 (1940) 193
[22] —— MARTIN, D. L. and ANGIER, R. P. *Trans. Amer. Soc. Met.* 31 (1943) 675
[23] MÜLLER, H. G. *Z. Metallk.* 31 (1939) 322
[24] KAISER, K. *Ibid* 19 (1927) 435
[25] BASSETT, W. H. and BRADLEY, J. C. *Trans. Amer. Inst. min. met. Engrs* 104 (1933) 181
[26] BALDWIN, W. M. *Ibid* 166 (1946) 591
[27] COOK, M. and RICHARDS, T. LL. *J. Inst. Met.* 73 (1946) 1
[28] —— —— *Ibid* 70 (1944) 159
[29] SACHS, G. and SPRETNAK, J. *Trans. Amer. Inst. min. met. Engrs* 140 (1940) 359
[30] BURGHOFF, H. L. and BOHLEN, E. C. *Ibid* 147 (1942) 144
[31] COOK, M. *J. Inst. Met.* 60 (1937) 159
[32] PALMER, E. W. and SMITH, C. S. *Trans. Amer. Inst. min. met. Engrs* 147 (1942) 164
[33] BALDWIN, W. M. and HOWALD, T. S. *Ibid* 166 (1946) 86
[34] COOK, M. and RICHARDS, T. LL. *J. Inst. Met.* 69 (1943) 201
[35] KOSTER, W. *Verband Mat. Pruf. Technik. Bericht* 1927, p 7
[36] BAUER, O., GOLER, F. v. and SACHS, G. *Z. Metallk.* 20 (1928) 202
[37] FAHRENHORST, W., MATTHAES, K. and SCHMID, E. *Z. Ver. dsch. Ing.* 76 (1932) 797

7

DIFFUSION OF METALS IN METALS

A. D. Le Claire

IT WAS early observed that if the partition separating two different gases or liquids was removed, changes occurred which finally led to a uniform concentration of each component on both sides of the plane of partition, and this tendency towards homogenization was termed diffusion. Although a number of phenomena involving diffusion in solids such as the carbonization of steel, the cold welding of metals,[1] and the making of alloys by heating metal powders[2] were familiar quite early on in the last century, it was not until 1896, when ROBERTS-AUSTEN[3] measured the rate of diffusion of gold in lead, that the first quantitative study of the subject for its own sake was made. Since then diffusion has been studied in a great number of systems and there has grown up, particularly in more recent years, an appreciation of the very important part played by diffusion processes in numerous metallurgical phenomena (see, for example, some of the reviews on the subject, particularly by MEHL[4–6]). For this reason, as well as the interest in the subject for its own sake as a branch of the wider field of the physics of the solid state, a thorough understanding of the factors affecting the rates of diffusion in solid metals is of great importance. Such an understanding can only be based upon a substantial foundation of accurate and careful measurements of diffusion coefficients in a large number of systems, and it is the main purpose of this chapter to describe in some detail a phenomenological theory in terms of which such measurements should be made and to review briefly the experimental methods which have been used. Finally we shall mention some of the more important results which have been obtained and the conclusions which have been drawn from them.

The first mathematical treatments embodied the realization, later to be modified as shall be seen, that for diffusion to take place between two points there must be a difference of concentration between them. Guided no doubt by the analagous case of thermal conductivity which had been studied just previously by FOURIER in 1854, FICK[7] defined a

diffusion coefficient D, for diffusion occurring along the x direction only, by the relation:

$$S = -AD \cdot \frac{\partial c}{\partial x} \quad \ldots\ (1)$$

where S is the amount of material diffusing in unit time through an area A, perpendicular to the x axis, at a point where the concentration gradient is $\partial c/\partial x$, the negative sign expressing the fact that diffusion takes place down the concentration gradient. D has the dimensions L^2T^{-1}. The form of equation 1 is justified by more fundamental treatments of the diffusion process, as will be seen later.

Equation 1 is applicable only to stationary flow. A more general form readily derived from it is:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D \frac{\partial c}{\partial x}\right) \quad \ldots\ (2)$$

where $\partial c/\partial t$ is the rate of accumulation of the diffusing material at a point where the concentration gradient is $\partial c/\partial x$, the concentration c being expressed in the same units on each side of the equation.

The extension of this to diffusion in three dimensions is readily made:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D_x \frac{\partial c}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_z \frac{\partial c}{\partial z}\right)$$

The three different diffusion coefficients D_x, D_y and D_z provide for a possible anisotropy in diffusion, which is in fact found in non-cubic crystals such as bismuth[8] and zinc.[9] Anisotropy has also been reported, most surprisingly, in cubic crystals,[10] but this has not yet been confirmed.

If we now make the assumption that D is independent of concentration, equation 2 becomes:

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad \ldots\ (3)$$

Until 1933 all experiments in metallic diffusion were interpreted in terms of solutions of this equation satisfying the boundary conditions appropriate to the particular experimental arrangement being used; or perhaps it would be more true to say that experimental arrangements were adopted for which solutions of equation 3 were available.

An approximate solution of equation 2 which does not involve the assumption that D is constant was given by BOLTZMANN[11] as early

as 1894. In 1933 MATANO[12] applied this to the diffusion data for the Ni-Cu system obtained about a year earlier by GRUBE and JEDELE[13] and showed that there is a marked, approximately tenfold, variation of D with concentration. Grube and Jedele themselves, and later Jedele[14] for the alloys of gold with platinum, palladium, and nickel, had pointed out that D was by no means constant for the systems studied by showing that their data could not be fitted to a solution of equation 3 over the whole range of concentration for a single value of D. Subsequently such a variation of D has been found in all alloy systems investigated by the Matano and Boltzmann or the Grube and Jedele method. Strictly speaking then, all diffusion measurements should be interpreted in terms of equation 2; all those employing equation 3 will give in general only a mean value of D over some range of concentration, and so the results will not be nearly so useful for theoretical purposes. Grube and Jedele, and Jedele did, however, calculate a series of values for D at different concentrations using equation 3, and their method, approximate though it is, has been used more recently.[15]

A variation of D with concentration calls for a reconsideration of Fick's equations 1 and 2, and we may look at the matter from the following point of view. An equation

$$\text{Mass flow} = \text{Constant } D \times \text{Gradient of concentration} \quad \ldots \ldots \quad (4)$$

has been postulated to describe diffusion processes, and as a first approximation D has been assumed constant. Experiment has shown, however, that D varies with concentration, and if we are to retain the general form of equation 4 we can either allow that D varies with concentration so that a complete phenomenological account of diffusion in any system may be expressed in terms of equation 4 and a relation expressing the variation of D with c, or slightly modify the equation to read:

$$\text{Mass flow} = \text{Constant } D_{02} \times \text{Gradient of some function of concentration}$$

and then set out to find a suitable form for the function of the concentration $f(c)$ so that D_{02} becomes a true constant. The distinction between these two modes of procedure seems first to have been made by ULLER in 1931.[16] We may then write our fundamental diffusion equation in one of three forms:

$$\left.\begin{aligned} a\ \text{Mass flow} &= D_{01}\ \frac{\partial c}{\partial x}, \quad \text{leading to } \frac{\partial c}{\partial t} = D\ \frac{\partial^2 c}{\partial x^2} \\ b\ \text{Mass flow} &= D(c)\ \frac{\partial c}{\partial x}, \quad \text{leading to } \frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D\ \frac{\partial c}{\partial x}\right) \\ c\ \text{Mass flow} &= D_{02}\ \frac{\partial \mathrm{f}(c)}{\partial x}, \text{ leading to } \frac{\partial c}{\partial t} = D_0\ \frac{\partial^2 \mathrm{f}(c)}{\partial x^2} \end{aligned}\right\} \quad \ldots . \quad (5)$$

Equations 5*b* and 5*c* are the only ones applicable to interdiffusing metal systems. Equation 5*a* has been applied but gives only an average value of D and is valid strictly only for self diffusing systems.

Clearly, if $D(c)\ .\ \mathrm{d}c$ is a perfect differential, then equations 5*b* and 5*c* are equivalent.

Some suggestions have been made recently[17,18] as to the form of $\mathrm{f}(c)$, and equation 5*c* incorporating these has been applied to experimental data on the diffusion of carbon in iron and of zinc in brass. The range of variation of the measured diffusion coefficient with concentration is much reduced, and in the case of carbon in iron calculated values of D_{02} are very nearly constant. But we shall see, when we have treated the diffusion process from a more fundamental and less postulatory basis, that there exist no *a priori* reasons for believing that D_{02} should be constant. It will be found that equation 5*b* is the most suitable form in which to represent diffusion processes and that $D(c)$ can be expressed as the product of two factors, one partly dependent upon the departure of the solid solution from ideality and proportional to the average diffusion force acting on one atom and the other a mobility term determining the average velocity produced by this diffusion force. Both of these terms will be expected to vary with concentration, and it will appear that only when certain assumptions are made about the dependence of the mobility upon concentration does $D(c)\ .\ \mathrm{d}c$ become a perfect differential and an equation of the form 5*c* hold. In other words, the validity of the suggestions about the possible forms of $\mathrm{f}(c)$ that have so far been made depends upon the truth of the assumptions they imply about the variation of mobility with concentration.

With equation 5*b* as the most appropriate one for the representation of diffusion data, the task of diffusion theory is to interpret the various quantities contained in the expressions we shall derive for $D(c)$ in terms of atomic properties of the diffusing atoms, or perhaps in the

first place to relate them to other measurable properties of the substances, *e.g.* melting point, sublimation energies *etc* which might appear the more intimately related to the diffusion process. Success in this direction will depend to a certain extent upon the correct interpretation of diffusion experiments and the separation of measured diffusion coefficients into their physically significant parts in the way which will now be described.

PHENOMENOLOGICAL THEORY

The free energy G of any phase consisting of a number of components is a function of the mol number c_i of each component i, of the temperature, the pressure, and of any other constraint such as a centrifugal or electric field which may be present. If the composition changes by amounts dc_1, dc_2 *etc* while the temperature, the pressure, and other constraints remain the same, then the change in free energy is:

$$dG = \frac{\partial G}{\partial c_1} dc_1 + \frac{\partial G}{\partial c_2} dc_2 + \ldots = \mu_1 dc_1 + \mu_2 dc_2 + \ldots$$

where the μ_i, the 'chemical potentials,' are functions of all the c_i and of the temperature pressure and of the other constraints. μ_i is then the increase in free energy per mol consequent upon the addition of an infinitesimal amount of the substance i to the phase under consideration.

Now the chemical potentials of the components of a system in which diffusion is taking place are quantities far more fundamental in determining the process of diffusion than are their respective concentrations, for a system is in thermodynamic equilibrium only when the chemical potential μ_i of each component i is the same at every point in each phase throughout the system. If the chemical potential of component i at one point is greater than at a second point, then, in those systems in which we are interested, flow by diffusion of the component i will take place from the first point to the second. We see then that the essential requisite for diffusion to occur between two points is that there exists between them a difference in chemical potential and that diffusion takes place down the chemical potential gradient which is not also necessarily the direction from higher to lower concentration. Qualitative demonstrations of this fact have been discussed by DARKEN,[17] who states that 'cementite in intimate contact with austenite saturated therewith exhibits no tendency to lose carbon to the austenite; in fact, austenite supersaturated with respect to cementite tends to

lose carbon to produce this higher carbon phase, that is, carbon diffuses from a region (austenite) where it is present at a lower concentration to a region (cementite) where it is present in a higher concentration'. The reason for this is that the chemical potential of carbon in cementite is less than it is in austenite in spite of the concentrations being in the reverse order of magnitude. Such examples of uphill diffusion (uphill against the concentration gradient) are not very common, but being extreme cases they illustrate best the importance of chemical potentials in determining diffusion processes.

Consider one gram atom of material to diffuse from a point O where the chemical potential is μ_O to a neighbouring point O' distance δx away, where the chemical potential is μ_O' ($\mu_O > \mu_O'$). Then the decrease in free energy per gram atom is given by:

$$\mu_O - \mu_O' = \frac{\partial \mu}{\partial x}\,\delta x + \left(\frac{\partial^2 \mu}{\partial x^2} \cdot \frac{(\delta x)^2}{2!}\right) + \ldots$$

and when second order terms can be ignored we may consider there to be an effective diffusion force acting on one gram atom given by $F = -\partial\mu/\partial x$. For the 'force' on one atom of type i we shall have:

$$f_i = -\frac{1}{N} \cdot \frac{\partial \mu_i}{\partial x} \qquad \ldots\ (6)$$

where N is Avogadro's number and μ_i is the chemical potential of the i^{th} component. This is a perfectly general result which we shall use as our starting point. The precise form of μ will depend upon the system being studied and upon the constraints placed upon it, such as stress, electric field, centrifugal field *etc* but we shall only consider in any detail a system in which μ is entirely determined by the temperature, pressure, and composition at each point. Generally too the systems will be isothermal and the applied pressure zero.

In ideal solid solution the chemical potential of the i^{th} component is given, in terms of its concentration, expressed as a mol fraction N_i by the relation:

$$\mu_i = k_0(T,p) + RT \log N_i \qquad \ldots\ (7)$$

but when the solution is not ideal the mol fraction is replaced by the activity a_i (referred to the pure i^{th} component as standard state), defined so that

$$\mu_i = k_0(T,p) + RT \log a_i \qquad \ldots\ (8)$$

where $k_0(T,p)$ is the free energy per mol of pure substance i and hence

a function of T and p only. Since we are dealing only with isothermal isobaric systems we can treat $k_0(T,p)$ as constant. The activity of pure i and for ideal solutions the ratio a_i/N_i ($= \gamma$, the 'activity coefficient') are both equal to unity; the departure of γ from this value is a measure of the departure of the solution from ideal. From equations 7 and 8, then, it might appear reasonable to postulate that the concentration in Fick's equation 1 should also be replaced by the activity so as to read:

$$\text{Mass flow} = D_{02}\frac{\partial a_i}{\partial x} \qquad \ldots . \quad (9)$$

or in other words to postulate that in equation 5c, $f(c) = a_i$.

This is just what DARKEN[17] and BIRCHENALL and MEHL[18] have done, and their work will be discussed in more detail later. As mentioned before, they hoped to obtain by this means a constant D_{02} in 9, but it will be seen that by substituting activity for concentration in this manner they were correctly allowing for the departure of the solution from ideality in so far as it affects the diffusion force acting on an atom but were implicitly making an assumption as yet unjustified about the dependence of mobility on concentration.

We shall now proceed to derive the general form of Fick's equation, starting from the principles embodied in relation 6. We need to consider two separate types of mechanism when discussing diffusion in metals.

1 The elementary atomic process determining the rate of diffusion involves the net movement of one atom only. This will be the case for 'hole' diffusion, diffusion of gases in metals, interstitial diffusion, and probably grain boundary diffusion. As pointed out by GLASSTONE, LAIDLER, and EYRING,[19] this does not necessarily mean that only one atomic species is involved in the diffusion but only that the diffusion process as observed is the net effect of a series of atomic processes each involving the movement of only one atom at a time.

2 Diffusion takes place by the mutual interchange of positions of two neighbouring atoms. For this process to lead to an observable change the participating atoms must, of course, be of different species.

Type 1. Mechanisms involving One Atom—The average diffusion force acting in the x direction on a single atom of species i is given by equation 6:

$$f_i = -\frac{1}{N}\cdot\frac{\partial \mu_i}{\partial x}$$

Let the mobility of the particle, that is, its average velocity under unit force, be G_i. Then if C_i is the average velocity of diffusion of the particle relative to same fixed lattice point we shall have:

$$C_i = -\frac{G_i}{\mathcal{N}} \cdot \frac{\partial \mu_i}{\partial x}$$

If n_i is the number of atoms of i per cc, the number Δn_i of atoms of i crossing unit area perpendicular to the x direction in unit time is given by:

$$\Delta n_i = -n_i \frac{G_i}{\mathcal{N}} \cdot \frac{\partial \mu_i}{\partial x} \qquad \ldots . \quad (10)$$

$$= -n_i \frac{G_i}{\mathcal{N}} RT \frac{\partial \log a_i}{\partial x} \quad \text{by equation 8}$$

$$= -n_i G_i kT \left(\frac{1}{\mathcal{N}_i} + \frac{\partial \log \gamma_i}{\partial \mathcal{N}_i}\right) \frac{\partial \mathcal{N}_i}{\partial x}$$

where a_i has been replaced by the activity coefficient γ_i ($a_i = \mathcal{N}_i \gamma_i$) and $\mathcal{N}_i$ is the mol fraction of the i^{th} component. The increase per unit time of the number of atoms per unit volume contained between planes at x and $x + \mathrm{d}x$ is then:

$$\frac{\partial n_i}{\partial t} = \frac{\partial}{\partial x}\left\{ n_i GkT \left(\frac{1}{\mathcal{N}_i} + \frac{\partial \log \gamma_i}{\partial \mathcal{N}_i}\right) \frac{\partial \mathcal{N}_i}{\partial x}\right\} \qquad \ldots . \quad (11)$$

Assuming that the atoms are of sufficiently the same size for Σn_i to be constant; equation 11 can be written, since $\mathcal{N}_i = n_i / \Sigma n_i$:

$$\frac{\partial \mathcal{N}_i}{\partial t} = \frac{\partial}{\partial x}\left\{ G_i kT \left(1 + \frac{\partial \log \gamma_i}{\partial \log \mathcal{N}_i}\right) \frac{\partial \mathcal{N}_i}{\partial x}\right\} \qquad \ldots . \quad (12)$$

so that the diffusion coefficient D'_i for a type 1 mechanism appears as:

$$D'_i = G_i kT \left(1 + \frac{\partial \log \gamma_i}{\partial \log \mathcal{N}_i}\right) \qquad \ldots . \quad (13)$$

This equation is identical with that derived by ONSAGER and FUOSS[20] for dilute electrolytes as far back as 1932, by DEHLINGER[21] for alloys, using a slightly different method of approach in 1936, and more recently, as a simple case of a more general relation, by LAMM.[22] The essential role played by activities and activity coefficients in the diffusion of liquids seems to have been realized for some considerable time but it is only more recently that their importance in metallic diffusion has been widely accepted.

Several interesting facts emerge from a study of equation 13. Since by the Gibbs-Duhem relation for a binary system we have:

$$\frac{\partial \log \gamma_1}{\partial \log N_1} = \frac{\partial \log \gamma_2}{\partial \log N_2} \quad \ldots . \quad (14)$$

the expressions for the rates of diffusion of the two components of a binary alloy will differ only in the mobility term G, and only when the mobilities of both components are equal will the diffusion coefficients be the same. Until quite recently the assumption usually seems to have been made in the interpretation of metallic diffusion experiments in binary systems that this equality always held, that is to say, that at any one concentration one diffusion coefficient alone was sufficient to express the rate of diffusion in any binary system, whatever the mechanism, although in many other studies involving diffusion in the solid state, such as scaling,[23] diffusion in ionic crystals[24] and tarnishing,[25] it is generally allowed that the various components can diffuse at different rates and no suggestion that such behaviour is anomalous is ever made. From the recent experiments of SMIGELSKAS and KIRKENDALL[26] it seems almost certain, at least in the case of zinc and copper in α brass, that the separate rates of diffusion in alloy systems also are not necessarily equal. These two workers laid fine molybdenum wires along the opposite surfaces of a rectangular bar of brass and then plated the whole with a thick layer of copper as illustrated in *Figure 1*. After heating for some time to allow diffusion to take place, the bar was sectioned and the distance between opposite brass/copper interfaces, as defined by the molybdenum wires, was measured and found to be considerably less than the original distance by an amount much greater than could be accounted for by any changes in density consequent upon the diffusion. The conclusion is then that the zinc must have diffused out of the brass much faster than the copper could diffuse in.

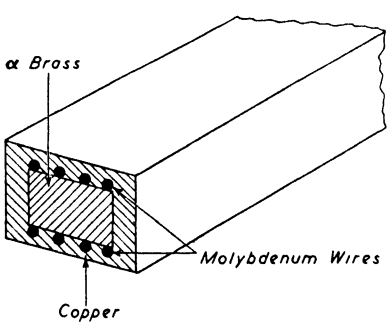


Figure 1. The experiment of Smigelskas and Kirkendall

Similar evidence of unequal diffusion rates of the two components of a binary alloy has been obtained for α aluminium bronzes.[27]

And again BUCKLE,[129] in attempting to locate the original interface of an interdiffusing couple of silver and zinc, found an anomalous volume change in the diffusion zone which he could only account for as due to the fact that the diffusion interchange across the α-β boundary was 'strongly to the advantage of the β phase'. That is to say there is a greater diffusion across the α-β interface of Zn from the β into the α phase than of Ag in the reverse direction.

We might mention too that JEDELE[14] as far back as 1933 put forward evidence that the rates of diffusion of the two components of the alloys of gold with palladium, nickel and platinum were not equal, but this cannot be taken as conclusive on account of his method of analysis, involving as it does the use of equation 3 only. Thus is revealed an entirely new aspect of the subject of metallic diffusion and it becomes important to investigate experimentally and theoretically this 'Kirkendall Effect' in as large a number of systems as possible and to develop new and refined techniques for doing so.

The above remarks depend in part upon the validity of the application of the Gibbs-Duhem relation 14 to phases in which the diffusion is taking place. We must inquire if the density and other physical properties at a point in the diffusion zone are the same as that for an ordinary homogeneous alloy of the same composition as at that point. If not, then the activity coefficients measured as they are for alloys in homogeneous equilibrium will not be the same necessarily as for an alloy in which diffusion is taking place. That such a change in density may occur is suggested by the experiments of Smigelskas and Kirkendall.[26] If confirmed then the relation 12 for D'_i, although correct as it stands, will only be approximately correct when measured values of activity coefficients are inserted, and furthermore D'_1 and D'_2 will not necessarily differ *only* in the mobility terms. When the diffusion mechanism is of type 2 the diffusion coefficients are essentially the same for both components and the Gibbs-Duhem relation will be valid at all points of the diffusion zone.

Type 2. Mechanisms involving Two Atoms—Since in this case two neighbouring atoms are involved at the same time in the elementary rate determining process, we must write for the average diffusion force acting on an atom of species A in a binary solid solution of A and B:

$$f_A = -\frac{1}{N}\left(\frac{\partial \mu_A}{\partial x} - \frac{\partial \mu_B}{\partial x}\right) \qquad \ldots\ (15)$$

where μ_A and μ_B are the chemical potentials of the two components. This follows at once from an extension of the argument leading up to equation 6. The force on the B atom exchanging position with the A atom is similarly given by:

$$f_B = -\frac{1}{\mathcal{N}}\left(\frac{\partial \mu_B}{\partial x} - \frac{\partial \mu_A}{\partial x}\right) = -f_A \quad \ldots . \quad (15a)$$

and is numerically equal to f_A, as expected. We have assumed in relation to 15 and 15*a* that the net changes in position of the two interchanging atoms are equal, which will be true at least in those cases where the atomic sizes do not differ appreciably.

We may now define a mobility term G_{AB} in the same way as for type 1, such that when unit force acts on an atom of A, G_{AB} is the average velocity with which A moves during the time taken for a complete A to B, B to A exchange process, which time includes that spent by the A atom as a neighbour of B before the exchange occurs. This latter proviso enables us later to relate G_{AB} to the relations for probability of exchange obtained from the various theoretical treatments of the diffusion process. The mobility of atoms of B under similar conditions may be defined as G_{BA}, but by the very nature of the process and the equality of f_A and f_B, G_{AB} and G_{BA} are equal and the subscripts may be written in either order.

When the force is given by equation 15, the average velocity of an atom of A during an A to B exchange is given by:

$$C_A = -\frac{G_{AB}}{\mathcal{N}}\left(\frac{\partial \mu_A}{\partial x} - \frac{\partial \mu_B}{\partial x}\right)$$

The number Δn_A of atoms of species A moving across unit area in unit time is equal to the number of atoms of A in volume C_A multiplied by the fraction of the neighbours of A which are on the average of type B, since only interchanges of A with B will lead to an observed change in the system. This fraction is clearly $\mathcal{N}_B$, the mol fraction of B. Exchanges of atoms of type A with others of the same type lead to no observable change in the concentration and so are of no importance in the diffusion. In accordance with this, the average force acting on an A atom during an A to A interchange is:

$$f_{AA} = -\frac{1}{\mathcal{N}}\left(\frac{\partial \mu_A}{\partial x} - \frac{\partial \mu_A}{\partial x}\right) = 0$$

We have now for Δn_A, when n_A is the number of atoms of A per unit volume:

$$\Delta n_A = - \mathcal{N}_B n_A \frac{G_{AB}}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial x} - \frac{\partial \mu_B}{\partial x} \right)$$

$$= - \mathcal{N}_B n_A \frac{G_{AB}}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial \mathcal{N}_A} - \frac{\partial \mu_B}{\partial \mathcal{N}_A} \right) \frac{\partial \mathcal{N}_A}{\partial x} \quad \ldots . \quad (16)$$

and for the rate of accumulation of A atoms per unit time, between planes at x and $x + dx$:

$$\frac{\partial n_A}{\partial t} = \frac{\partial}{\partial x} \left\{ \frac{\mathcal{N}_B n_A G_{AB}}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial \mathcal{N}_A} - \frac{\partial \mu_B}{\partial \mathcal{N}_A} \right) \frac{\partial \mathcal{N}_A}{\partial x} \right\}$$

$$\text{or} \quad \frac{\partial \mathcal{N}_A}{\partial t} = \frac{\partial}{\partial x} \left\{ \frac{\mathcal{N}_A \mathcal{N}_B G_{AB}}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial \mathcal{N}_A} - \frac{\partial \mu_B}{\partial \mathcal{N}_A} \right) \frac{\partial \mathcal{N}_A}{\partial x} \right\} \quad \ldots . \quad (17)$$

if we can assume, as before, that the atomic radii of A and B are sufficiently the same for us to put $n_A + n_B = \text{constant}$.

From equation 17 we see that the diffusion coefficient D''_A for a type 2 mechanism is given by:

$$D''_A = \frac{\mathcal{N}_A \mathcal{N}_B G_{AB}}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial \mathcal{N}_A} - \frac{\partial \mu_B}{\partial \mathcal{N}_A} \right)$$

$$= \mathcal{N}_A \mathcal{N}_B G_{AB} kT \left(\frac{\partial \log \mathcal{N}_A \gamma_A}{\partial \mathcal{N}_A} - \frac{\partial \log \mathcal{N}_B \gamma_B}{\partial \mathcal{N}_A} \right) \quad \ldots . \quad (18)$$

$$= G_{AB} kT \left\{ 1 + \mathcal{N}_A \mathcal{N}_B \left(\frac{\partial \log \gamma_A}{\partial \mathcal{N}_A} - \frac{\partial \log \gamma_B}{\partial \mathcal{N}_A} \right) \right\} \quad \ldots . \quad (18a)$$

since $d\mathcal{N}_A = - d\mathcal{N}_B$. Equation 18*a* was first derived by Dehlinger[21] in 1936 by a slightly different method. By employing the Gibbs-Duhem relation 14, which as we have seen should be exactly applicable to the changing concentrations in diffusion zones where a type 2 mechanism is operating, equation 18*a* reduces for binary systems to a form identical with that obtained for type 1 mechanisms. By equation 14:

$$\frac{\partial \log \gamma_B}{\partial \mathcal{N}_B} = \frac{\mathcal{N}_A}{\mathcal{N}_B} \times \frac{\partial \log \gamma_A}{\partial \mathcal{N}_A} = - \frac{\partial \log \gamma_B}{\partial \mathcal{N}_A}$$

Therefore

$$D''_A = G_{AB} kT \left\{ 1 + \mathcal{N}_A (1 - \mathcal{N}_A) \frac{\partial \log \gamma_A}{\partial \mathcal{N}_A} + \mathcal{N}_A{}^2 \frac{\partial \log \gamma_A}{\partial \mathcal{N}_A} \right\}$$

$$= G_{AB} kT \left\{ 1 + \frac{\partial \log \gamma_A}{\partial \log \mathcal{N}_A} \right\} \quad \ldots . \quad (19)$$

which is identical in form with equation 13. The condition that $D_A = D_B$, required for a type 2 mechanism, is clearly satisfied on account of the symmetry of equation 18*a* and the equality of G_{AB} and

G_{BA}. STEARN, IRISH, and EYRING[28] have derived a relation similar to equation 19 in which the bracketed term is

$$\left(1 + \frac{\partial \log \gamma_A}{N_B \partial \log N_A}\right)$$

but in their derivation appear to have ignored the change in the entropy of mixing of B which necessarily accompanies the change in that of A in this type 2 mechanism. This is allowed for in the above method of derivation.

BECKER[29] has given an expression for the diffusion coefficient which, while equivalent to the expressions 13 or 19 which we have derived, is in a form more convenient for some theoretical purposes[29, 30] since it involves the free energy directly instead of the less tangible but more readily measurable activity coefficient. Becker's expression is, with change of notation:

$$D_A = N_A N_B G_{AB} \frac{\partial^2 G}{\partial N_A^2} \quad \ldots . \quad (20)$$

where G is the total free energy of the solution divided by the total number of atoms present. Now

$$\mathcal{N} \partial G = \mu_A \mathrm{d}N_A + \mu_B \mathrm{d}N_B = (\mu_A - \mu_B)\, \mathrm{d}N_A$$

where $\mathcal{N}$ = Avogadro's number, and

$$\frac{\partial^2 G}{\partial N_A^2} = \frac{1}{\mathcal{N}} \left(\frac{\partial \mu_A}{\partial N_A} - \frac{\partial \mu_B}{\partial N_A}\right)$$

Inserting this value of $\partial^2 G / \partial N_A^2$ into equation 20, we obtain after substitution for μ an expression identical with 18 which as we have already seen reduces to equation 19.

Several Mechanisms Operating—When more than one mechanism is operative, as for example when some atoms are diffusing by a hole mechanism, as discussed later, and others by an exchange process, then we shall have to define a mobility for each mechanism and our equation will take the form:

$$D = (G_1 kT + G_2 kT) \left(1 + \frac{\partial \log \gamma}{\partial \log \mathcal{N}}\right)$$

However, in the majority of cases G_1 and G_2 will be very different, so that the smaller of them may be ignored and we arrive again at equation 13 or 19.

Of more interest is the case where diffusion is occurring by some

mechanism through the lattice and also along the grain boundaries. We may now perhaps write D in the form:

$$D = \{x_1 G_1 kT + (1 - x)\, G_2 kT\} \left(1 - \frac{\partial \log \gamma}{\partial \log N}\right)$$

where x is the effective fraction of a unit area through which diffusion is occurring by the grain boundary process, although the exact evaluation of x in any particular case would be very difficult. This point has been discussed by Mehl.[6] We need to know whether there is a sharp boundary where the volume or lattice diffusion gives place to the more rapid grain boundary diffusion and, if so, its depth below the surface of the grain, or whether the transition is a more gradual one so that as we move from the surface into the grain the observed diffusion coefficient changes continuously. Practically nothing is known about this.

Uphill Diffusion—Equations 12 or 19 and 20 indicate very general conditions for uphill diffusion to occur, for whatever the mechanism of diffusion in a binary system, D will be negative whenever

$$\frac{\partial \log \gamma_A}{\partial \log N_A} + 1 < 0 \qquad \ldots\ldots (21)$$

or when (Becker)

$$\frac{\partial^2 G}{\partial N_A^{\,2}} < 0 \qquad \ldots\ldots (21a)$$

Also, when either of these expressions is equal to zero the diffusion coefficient will be zero and no diffusion will take place however large the concentration gradient. It would be most interesting to perform experiments on systems for which the activity is known to have a maximum or minimum value, as well as to correlate cases of uphill diffusion with relation 21. No work along these lines has yet been reported, nor indeed has any experimental work been performed, until very recently by DANIEL,[30] on systems exhibiting uphill diffusion. The reason for this probably is, as Miss Daniel points out, the difficulty of isolating the uphill diffusion process from other phenomena such as nucleus formation, grain growth and precipitation, with which it is so often associated. By employing the quite unusual properties of the alloy Cu_4FeNi_3, Miss Daniel has been able to avoid this difficulty and to study the process of uphill diffusion in this alloy. Her results provide a confirmation of equation 20.

Conditions 21 only apply to systems for which the chemical potential

is a function of temperatures and concentration alone. A more general form of the condition for uphill diffusion in all binary systems is clearly:

$$\frac{\partial \mu_A}{\partial N_A} < 0 \qquad \ldots . \quad (21b)$$

which is equivalent to Becker's relation 21*a*. μ_A may now include energy terms due to strain, electrical fields, centrifugal fields *etc.* Condition 21*b* may then be satisfied by the presence of suitable distributions of stress for example, and KONOBEEVSKY[31] has discussed cases where uphill diffusion arising from this cause might occur.

We may digress a moment to consider the physical meaning of uphill diffusion. In an alloy AB, the lattice energy of A atoms may roughly be regarded as due partly to A—A and partly to B—A bonds, and the strengths and therefore the energies associated with these two types of bond will in general be different; in ideal solution they are equal. When atoms of A diffuse normally, that is in such a direction that the concentration of A tends to become uniform, there is an increase in entropy and therefore a lowering of the free energy. At the same time there is a change, either an increase or decrease, of lattice energy since diffusion takes place between points of different concentration and therefore the relative numbers of A—A and B—A bonds per A atom change. Only when the increase in lattice energy is numerically less than the decrease in the entropy term of the free energy will diffusion be normal. In such cases the total free energy is a minimum for a single homogeneous phase. Uphill diffusion will occur when the increase of the entropy term in the free energy consequent upon the diffusion of A from a point of lower to one of higher concentration is less than the concomitant decrease in lattice energy. This process very frequently leads eventually to the formation of a two-phase structure, one rich in A and the other rich in B, for which evidently the free energy is a minimum. Clearly uphill diffusion requires that A—B bonds be much weaker than either A—A or B—B bonds, and also will be less likely to occur the higher we raise the temperature.[29]

In cases where the solid solution is ideal, expressions 13 and 19 both reduce to the well known Einstein form:[32]

$$D = GkT = \frac{GRT}{\mathcal{N}}$$

Even with ideal solutions the mobility will in general vary with the composition, although we might expect the extent of the variation to be small since in an ideal solution the fields of force surrounding solvent and solute atoms are identical, at least in their equilibrium positions. When the mobility is quite independent of concentration, and the solution ideal, as for example with mixtures of the stable and radioactive isotopes of an element, equation 12 reverts to the usual Fick form:

$$\frac{\partial N_i}{\partial t} = \frac{D\partial^2 N_i}{\partial x^2}$$

where D is a constant.

Fick's equation in this form, the one so frequently used, is then strictly valid only for thermodynamically ideal solid solutions in which the mobility is constant.[33]

In non-ideal solutions the term $(1 + \partial \log \gamma / \partial \log N)$ is by no means a minor correction to diffusion measurements. For example, in the case of zinc in brass at 750°C, it ranges from 1·0 at zero zinc concentration to about 2·4 at 26 per cent zinc (data from reference 18). Also for the nickel-gold system Dehlinger,[21] using admittedly a very rough method for estimating $(1+\partial \log \gamma/\partial \log N)$, found that at 50 atomic per cent nickel this term was as low as 0·04 and that GkT at this concentration was very appreciably greater than the mean calculated from the measured values at small concentrations of gold and of nickel, while the reverse was true of the observed diffusion coefficient.

Interpretation of Diffusion Coefficients—We have now demonstrated that, whatever the mechanism of diffusion, the diffusion coefficient in binary systems is the product of two factors, one a mobility term GkT which as will be shown later (see equation 40) measures the extent to which atoms migrate in the lattice even in the absence of a concentration gradient, and the other a term expressing the departure of the solution from ideality, which, when multiplied by the concentration gradient, is proportional to the so-called diffusion force acting on an atom. Observed variations of the diffusion coefficient with concentration are then to be analysed into the variations with concentration of each of these two terms. Fortunately we can measure the variation of the second term by determination of activity in solid solution and so calculate directly from measured diffusion coefficients the variation of mobility with concentration. The ultimate problem in diffusion theory is then the interpretation of the activity and particularly the mobility values,

and their variation with temperature and concentration in terms of atomic quantities. The factors affecting the activity are of course independent of the mechanism of diffusion, the activity being an equilibrium property of a solution, and besides, the term involving the activity is the same for both type 1 and 2 mechanisms in binary systems. This is not so for the mobility. We might expect the mobility term, and its dependence on concentration for a type 1 mechanism, involving one atom, to be of a form quite different from that for a type 2 mechanism, involving two atoms, when these are expressed in terms of quantities related to the participating atom or atoms. A type 1 mechanism will be indicated whenever we can obtain evidence that D_A and D_B are unequal, as for example in an experiment of the Smigelskas and Kirkendall type as quoted earlier; but when D_A and D_B are not unequal it might be possible by considering this difference in form of the mobility terms to decide in any particular case which type of mechanism was responsible for the observed diffusion. A more detailed atomic theory employing models of the diffusion process will give numerical estimates of the mobility and so, by comparison of these with observed values, enable a choice to be made of the appropriate atomic mechanism, but at this stage we are only concerned in employing any possible difference in form of the mobility terms to help in the broader classification into type 1 or type 2 mechanism.

We shall show later that for ternary and higher systems the situation is more complicated because the terms in the diffusion coefficient involving the activity are quite different for the two types of mechanism.

Two papers have appeared recently by BIRCHENALL and MEHL[18] and FISHER, HOLLOMAN, and TURNBULL[34] in which assumptions are made in effect, as to the different forms of dependence of the mobility term upon concentration for the two mechanisms, and in the former paper conclusions have been drawn, in line with the above remarks, about the nature of the mechanism of diffusion in the two cases of carbon diffusing in iron and of zinc in α brass.

By starting with their assumption that the fundamental equation for diffusion should be written:

$$\text{Mass flow} = D\frac{\partial a}{\partial x} \qquad \ldots . \quad (9)$$

where D is supposedly independent of concentration, Birchenall and

Mehl readily show that the observed diffusion coefficient $D(c)$ is related to D_1 by the equation:

$$D(c) = D_1\left(\gamma_1 + \frac{\partial\gamma_1}{\partial \log N_1}\right) \quad \ldots\ (22)$$

where D_1 is the value of D for a type 1 mechanism.

Equating this to our own expressions for $D(c)$ for a type 1 mechanism from equation 13, we have:

$$GkT\left(1 + \frac{\partial \log \gamma_1}{\partial \log N_1}\right) = D_1\left(\gamma_1 + \frac{\partial\gamma_1}{\partial \log N_1}\right)$$

from which we see that equation 9 involves the assumption that the mobility term GkT is proportional to the activity coefficient only, but that it does correct in the right way for the departure of the solution from ideality.

Values of D_1 are then calculated from the observed coefficient of diffusion of carbon in iron, using activity data taken from the paper by SMITH[110] and measurements on the diffusion of carbon in iron by WELLS and MEHL.[118] From structural considerations this is almost certainly a type 1 mechanism (interstitial diffusion), and we find that D_1 is indeed remarkably constant over a considerable range of concentration. We have then some justification for supposing that for a type 1 mechanism

$$G_1kT = D_1\gamma \quad \ldots\ (23)$$

so that the two factors entering into the diffusion coefficient $D(c)$ in equation 5*b* may be combined into a single factor f(c) and the diffusion equation written in the form of 5*c* or 9.

DARKEN[35] has already pointed out that relation 23 is implied in equation 9 and has criticized it on the grounds that the activity is an equilibrium property of a solid solution and therefore should not be expected to enter into an expression for such a kinetic quantity as the mobility. But the factors which determine the magnitude of the mobility will also be those which fix the value of the activity, namely the energies of nearest neighbour interactions, and there seems to be no reason to object to equation 23 on these grounds alone. It is, however, at the moment a purely experimental relation, and until justified theoretically it is safer to employ the equation 5*b* and not 5*c* for the representation of diffusion phenomena since it may not be valid for all type 1 mechanisms. In particular, as shown earlier, it is not valid for the

diffusion of zinc in α brass which the experiments of Smigelskas and Kirkendall show to be by a type 1 mechanism.

Birchenall and Mehl then arrive at an expression for a type 2 mechanism diffusion constant by supposing that if the term $(\gamma_1+N_1\partial\gamma_1/\partial N_1)$ in equation 22 be regarded 'as a probability factor expressing the fraction of a given number of atoms that will jump under existing concentrations and activity conditions, the overall probability when two different atoms must jump together will be the product of two such probability factors, one for each component'. Therefore:

$$D(c) = D_{12}\left(\gamma_1 + N_1\frac{\partial\gamma_1}{\partial N_1}\right)\left(\gamma_2 + N_2\frac{\partial\gamma_2}{\partial N_2}\right) \quad . \ . \ . \ . \quad (24)$$

an equation which, as these authors point out, is symmetrical with respect to both components, as any equation representing a type 2 mechanism must be. We should not regard the term $(\gamma_1+N_1\partial\gamma_1/\partial N_1)$ as expressing a probability of jumping, for that is entirely contained in our mobility terms, so that we cannot accept the derivation or result of equation 24. However, we see that it reduces by means of equation 14 to:

$$D(c) = D_{12}\gamma_1\gamma_2\left(1 + \frac{\partial \log \gamma_1}{\partial \log N_1}\right)^2 \quad . \ . \ . \ . \quad (25)$$

so that

$$GkT = D_{12}\gamma_1\gamma_2\left(1 + \frac{\partial \log \gamma_1}{\partial \log N_1}\right) \quad . \ . \ . \ . \quad (25a)$$

Ignoring the fact that the bracketed term in equation 25 is squared, it corresponds to our equation 19 if we assume that the mobility for a type 2 mechanism is proportional only to the product of the activity coefficients of the two components

$$GkT = D'_{12}\gamma_1\gamma_2 \quad . \ . \ . \ . \quad (26)$$

which might appear quite reasonable in view of the success of equation 23 in dealing with a case having almost certainly a type 1 mechanism. Equation 26 is equivalent also to the assumption made by Fisher, Holloman, and Turnbull[34] in deriving, on the basis of the reaction rate theory,[19] an expression for a type 2 diffusion mechanism. They obtain the relation:

$$D = D_0\left\{N_A\gamma_A\left(\gamma_B + N_B\frac{\partial\gamma_B}{\partial N_B}\right) + N_B\gamma_B\left(\gamma_A + N_A\frac{\partial\gamma_A}{\partial N_A}\right)\right\}$$

which reduces simply to:

$$D = D_0\gamma_A\gamma_B\left(1 + \frac{\partial \log \gamma_A}{\partial \log N_A}\right) \quad . \ . \ . \ . \quad (27)$$

We have seen that D_1 in equation 23 is constant for at least one example of a type 1 mechanism. D'_{12} in equation 26 has been calculated by Fisher *et al.*, and D_1 in equation 23 and D_{12} in equation 25 by Birchenall and Mehl, for the diffusion of zinc in α brass, using diffusion data selected from the paper by Rhines and Mehl.[73] None of these turned out to be constant, although the range of variation was found to decrease in the order D_1 (equation 23), D'_{12} (equation 26), D_{12} (equation 25).

There is nothing profoundly significant in this. Least of all are we justified in assuming, as Birchenall and Mehl do, that since D_{12} is more nearly constant in this case than is D_1, then the mechanism responsible for the diffusion of zinc in α brass is of type 2. The results merely indicate that none of the equations 23, 25*a*, or 26 are true expression for the dependence upon concentration of the mobility of zinc in α brass.

If we wish to make use of the possible difference in form of the concentration dependence of the mobilities for determining which type of mechanism is operative in any particular case, we must discover this difference by theoretical means or by studying a number of systems in which the type of mechanism is definitely known.

TERNARY AND HIGHER SYSTEMS

The above treatment has been principally for binary systems. For ternary and higher systems, equation 13 for type 1 mechanisms remains unchanged:

$$D_i = G_i kT\left(1 + \frac{\partial \log \gamma_i}{\partial \log N_i}\right)$$

although the addition of the third or further components will of course alter the values of the mobilities and activities of the original components. Furthermore, the differences between the diffusion coefficients of the various components are no longer due, necessarily, only to differences in the mobility term, since now

$$\left(1 + \frac{\partial \log \gamma_i}{\partial \log N_i}\right) \neq \left(1 + \frac{\partial \log \gamma_j}{\partial \log N_j}\right)$$

as was the case for a two-component system when $i = 1$ and $j = 2$.

For type 2 mechanisms we must consider, in following the diffusion of say A atoms, the possibilities of migration of A by exchanges with atoms of each of the other species B, C, D *etc.* Each such exchange

process will have in general a different mobility, and we may designate these G_{AB}, G_{AC}, G_{AD} *etc.* The observed flow of A due to exchanges of A with B (say) is then still given by equation 18, but the equation must be left in the form:

$$(D''_A)_{AB} = G_{AB}kTN_AN_B\left(\frac{\partial \log N_A\gamma_A}{\partial N_A} - \frac{\partial \log N_B\gamma_B}{\partial N_A}\right)$$

since we can no longer write $dN_A = -dN_B$ in general. Here $(D''_A)_{AB}$ is the diffusion coefficient of A for AB exchanges only. When all possible exchange processes can occur, the observed diffusion coefficient of A is given by:

$$D''_A = \sum_{X = B, C\ etc} G_{AX}kTN_AN_X\left(\frac{\partial \log N_A\gamma_A}{\partial N_A} - \frac{\partial \log N_X\gamma_X}{\partial N_A}\right) \quad . \quad . \quad . \quad . \quad (28)$$

with similar expression for D''_B *etc.*

This does not reduce to the form of equation 19, so that in ternary and higher systems the form of the phenomenological equation for the diffusion coefficient depends upon the type of mechanism considered. There would appear then to be a possibility of distinguishing between type 1 and type 2 mechanisms on phenomenological grounds only, by studying such systems. Unfortunately the use of equations 19 and 28 involves knowledge of the activities of the components of ternary alloys, and at the present time very little such information is available. We are also faced with the same difficulty as before over the nature of the concentration dependence of G_{AX}.

However, in the experiments of JOHNSON[36] as interpreted by DARKEN[37] and LE CLAIRE,[38] we have an example of how, by combining observations on a binary system with observations on a particularly simple ternary system, these difficulties can be overcome. Johnson first measured the diffusion coefficient of silver and gold in a 50 atomic per cent alloy of these two metals. If the mechanism is of type 2, we shall have:

$$D''_{AgAu} = G_{AgAu}kT\left(1 + \frac{\partial \log \gamma_{Ag}}{\partial \log N_{Ag}}\right) \quad . \quad . \quad . \quad . \quad (29)$$

but if of type 1, then, as we shall show later, the observed diffusion coefficient is given by:

$$D'_{AgAu} = \tfrac{1}{2}kT(G_{Ag} + G_{Au})\left(1 + \frac{\partial \log \gamma_{Ag}}{\partial \log N_{Ag}}\right) . \quad . \quad . \quad . \quad (29a)$$

since $N_{Ag} = N_{Au} = \frac{1}{2}$. He then measured the self diffusion coefficients of radioactive silver and of radioactive gold in an alloy of the same

composition (50 per cent). These are in effect studies of the diffusion of one component, say radioactive gold in a ternary alloy of radioactive gold, stable gold and silver in which the *total* concentration of gold and of silver is everywhere the same and the only concentration gradient that of radioactive gold. Since chemically the system is quite homogeneous, when the difference in mass between radioactive and stable isotopes is ignored, the activity coefficients are everywhere constant, the activity coefficient of an element being independent of the amounts of radioactive components. Equation 28 then reduces, in the case of the self diffusion of gold, to:

$$D''_{\mathrm{Au}} = G_{\mathrm{AuAu^*}}kT(\mathcal{N}_{\mathrm{Au}} + \mathcal{N}_{\mathrm{Au^*}}) + G_{\mathrm{AuAg}}kT\mathcal{N}_{\mathrm{Ag}} \quad . \quad . \quad . \quad . \quad (30)$$

since $\mathrm{d}\mathcal{N}_{\mathrm{Au^*}} = -\mathrm{d}\mathcal{N}_{\mathrm{Au}}$, where Au* denotes radioactive gold, with a similar expression for the self diffusion coefficient of silver. Now in Johnson's experiment $\mathrm{N}_{\mathrm{Ag}} = (\mathcal{N}_{\mathrm{Au}} + \mathcal{N}_{\mathrm{Au^*}}) = \frac{1}{2}$, so that equation 30 becomes:

$$D''_{\mathrm{Au^*}} = \tfrac{1}{2}kT(G_{\mathrm{AuAu^*}} + G_{\mathrm{Au^*Ag}}) \quad . \quad . \quad . \quad . \quad (31)$$

If, on the other hand, it is a type 1 mechanism with which we are dealing, we shall have:

$$D'_{\mathrm{Au^*}} = G_{\mathrm{Au^*}}kT \text{ and } D'_{\mathrm{Ag^*}} = G_{\mathrm{Ag^*}}kT \quad . \quad . \quad . \quad . \quad (32)$$

The various mobility terms $G_{\mathrm{Au^*}}$, $G_{\mathrm{Au^*Ag}}$ *etc* all refer to the same concentration of alloy (50 per cent) and so can be eliminated by simultaneous solution of the above equations. Since we can assume $G_{\mathrm{Au^*Ag}} = G_{\mathrm{AuAg}}$ *etc* we can use equations 29*a* and 32 to calculate D'_{AgAu} in terms of measured values of $D'_{\mathrm{Au^*}}$, $D'_{\mathrm{Ag^*}}$, and γ_{Ag}, and compare the result with the observed value. We can also calculate the corresponding quantity D''_{AgAu} for an assumed type 2 mechanism from equations 29 and 31 if we assume that $G_{\mathrm{Au^*Au}}$ can be neglected by comparison with $G_{\mathrm{Au^*Ag}}$. This is justifiable on the evidence that the self diffusion of Au in Au is very much less (approximately 20 times) than the self diffusion of gold in the alloy. The calculated values of D'_{AuAg} and D''_{AuAg} are then given by:

$$D'_{\mathrm{AuAg}} = \tfrac{1}{2}(D'_{\mathrm{Au^*}} + D'_{\mathrm{Ag^*}})\left(1 + \frac{\partial \log \gamma_{\mathrm{Ag}}}{\partial \log \mathcal{N}_{\mathrm{Ag}}}\right) \text{ at } \mathcal{N}_{\mathrm{Ag}} = \tfrac{1}{2} \quad . \quad . \quad . \quad . \quad (33)$$

$$D''_{\mathrm{AuAg}} = 2D''_{\mathrm{Au^*}}\left(1 + \frac{\partial \log \gamma_{\mathrm{Ag}}}{\partial \log \mathcal{N}_{\mathrm{Ag}}}\right) \text{ at } \mathcal{N}_{\mathrm{Ag}} = \tfrac{1}{2}. \quad . \quad . \quad . \quad (34)$$

The observed values of D_{AuAg} are compared with the values calculated from equation 33 and 34 in *Table I*,[38] the activity values being taken

from a paper by WAGNER and ENGELHARDT.[39] Although the agreement between observed and calculated values is slightly better for values calculated from equation 34, it is not sufficiently so for us to assert definitely that a type 2 mechanism is operative. The near correspondence for the two calculated values of D_{AuAg} arises because in this case, where N_{Au} and N_{Ag} are both $\frac{1}{2}$, it so happens that

$$2D_{Au^*} \simeq N_{Ag}D_{Au^*} + N_{Au}D_{Ag^*}$$

For a different alloy composition, the two calculated values of D_{AuAg}

Table I. Interpretation of Johnson's Experiments

Temperature °C	800	900	1000
D_{AuAg}, *observed*	$4{\cdot}5 \times 10^{-10}$	$24{\cdot}0 \times 10^{-10}$	$9{\cdot}7 \times 10^{-9}$
D'_{AuAg}, *calculated* (*eq.* 33)	$3{\cdot}8 \times 10^{-10}$	$21{\cdot}0 \times 10^{-10}$	$9{\cdot}3 \times 10^{-9}$
D''_{AuAg}, *calculated* (*eq.* 34)	$4{\cdot}5 \times 10^{-10}$	$23{\cdot}5 \times 10^{-10}$	$10{\cdot}45 \times 10^{-9}$

might be expected to differ by a much greater amount, so that Johnson's experiments repeated for a gold-silver alloy of composition very different from $N_{Ag} = N_{Au} = \frac{1}{2}$ might enable a more definite decision to be made on the mechanism of diffusion. Much valuable information could be derived from experiments conducted along these lines on a whole series of alloys.

It is interesting to note Johnson's experiments were only satisfactorily interpreted when the full expressions 29 and 29*a* for the diffusion coefficient were employed, which again illustrates the great importance of taking into account the departure of actual solid solution from the ideal in the evaluation of diffusion phenomena.

A more recent attempt by SEITZ[128] to explain Johnson's experiments does not distinguish between the two components of the diffusion coefficient and so requires quite a detailed and complex model. We have been able to do the same on a purely thermodynamic basis without recourse to any particular model.

Diffusion in ternary systems has not been studied to any very great extent and most investigations have been on the effect on the rates of diffusion of the components of a binary system of the addition of small quantities of a third component. In many cases the effects are considerable. Thus SMOLUCHOWSKI[40,41] finds that the rate of diffusion of carbon in γ iron (at one atomic per cent of carbon) is doubled by

the addition of 4 per cent of cobalt and halved by the addition of about 3 per cent of molybdenum or of only 1 per cent of tungsten (see *Figure 2*).

Manganese and nickel have practically no effect.[42] GRUBE and JEDELE[13] found also that the addition of only 0·5 per cent of manganese

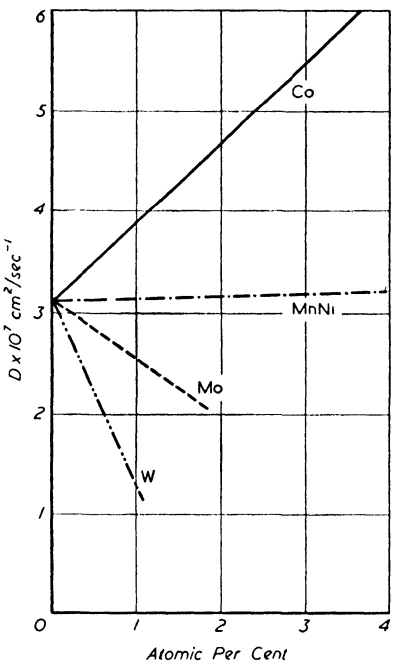


Figure 2. Diffusion coefficient of 1 per cent of carbon in iron in the presence of various amounts of cobalt, manganese, nickel, molybdenum, and tungsten[41]

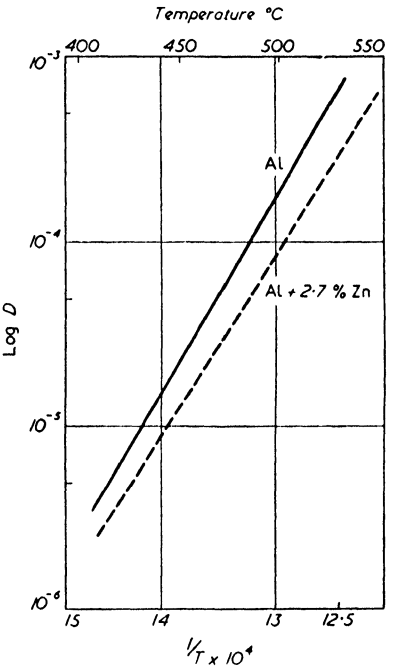


Figure 3. Diffusion of magnesium in pure aluminium and in an alloy of aluminium and 2·7 per cent zinc[43]

reduced by a factor of three the rate of diffusion of copper in nickel, and BUNGARDT and BOLLENRATH[43] that the diffusion coefficient of magnesium in aluminium was halved by the addition of 2·7 per cent of zinc (see *Figure 3*). The influence of tin, nickel, and lead on the diffusion of zinc in α brass has recently been studied by GERZRICKEN, GELLER, and TROFIMENKO.[44] While nickel has no influence, 2 per cent of tin increases the coefficient of diffusion ten times and 3·5 per cent of lead increases it about fifteen times. Such examples as these demonstrate the great importance of using very pure materials in experimental work on diffusion if the results are to be of any theoretical

importance. It is quite possible that even very minute quantities of impurity may have in some cases a very considerable effect upon rates of diffusion in metals by associating with those imperfections in the lattice (dislocations, holes, mosaic boundaries *etc*) whose presence is often considered responsible for solid diffusion (see page 336). We are reminded here of the great influence of impurities in the diffusion and mobility of gaseous ions.[45] The discrepancies between the observed values of diffusion coefficients measured by different observers for the same system might be traceable in part at least to the effects of impurities.

There are many cases of industrial importance where it may be useful to accelerate or decrease the rate of diffusion of some component of a binary alloy, such as the diffusion of carbon into iron during the cementation process, in which the desired effect may be achieved by the addition of quite small quantities of some third element. In this connection HOUDREMONT and SCHRADER[46] have investigated the effect of the addition of small quantities of approximately one per cent of various elements such as copper, titanium, and aluminium, on the rate of diffusion of carbon into iron. HENSEL[47] has also shown that the diffusion of silver into thallium in the manufacture of surface layers of antifriction silver-thallium alloys is much assisted by the addition of a little indium.

From the theoretical viewpoint the addition of the third element to a binary system will affect not only the mobility of the two original components but also their activities. Practically no information is available on the activities of the components of ternary or higher alloys, so that it is impossible at present to assess whether it is the changes in the mobility or in the activity consequent upon the addition of a third component which are the more responsible for the observed changes in D. It is most unlikely that there will be no effect upon the mobility term, so that experiments on the influence of the addition of a third component on the diffusion of the components of a binary alloy may possibly be of assistance in elucidating the nature of the atomic processes producing diffusion. This viewpoint is also expressed in a recent paper by GERZRICKEN and DEGHTYAR,[48] in which is discussed the results of experiments on the influence of magnesium and antimony on the rates of diffusion of mercury from an alloy of mercury and lead.

We have not yet mentioned the conditions for uphill diffusion in ternary systems. For a type 1 mechanism, where the phenomenological

form of the diffusion coefficient is the same for ternary as for binary alloys, the condition for uphill diffusion is also the same, namely that given by equations 19 and 20:

$$1 + \frac{\partial \log a_A}{\partial N_A} < 0 \quad \text{or} \quad \frac{\partial \mu_A}{\partial N_A} < 0$$

For a type 2 mechanism however, the condition, which is considerably more complex and cannot be expressed purely in terms of thermodynamic quantities, is

$$\sum_{X=B,C,D\ etc} G_{AX} N_A N_X \left(\frac{\partial \log a_A}{\partial N_A} - \frac{\partial \log a_X}{\partial N_X} \right) < 0 \quad . \quad . \quad . \quad . \quad (35)$$

for uphill diffusion of A, and is seen to involve all the mobility terms. Cases might clearly arise in which an uphill diffusion of A by exchange of A with S is more than countered by a normal diffusion of A by exchange with T so that the observed diffusion process will appear normal.

The fact that the particular conditions for uphill diffusion in ternary systems depend upon the type of diffusion mechanism in operation might be employed in the determination of the nature of the mechanism in any particular case. We can imagine the possibility of finding, after investigation of a number of cases of uphill diffusion in ternary systems, that some are very well accounted for by means of equation 19, while others are definitely not. We should then be led to assume that in the former cases a type 1 mechanism was operative, and in the latter cases a type 2. This method of distinguishing between the two mechanisms cannot be employed in binary systems, for there the same condition for uphill diffusion applies for both mechanisms.

DERIVATION OF IDEAL DIFFUSION COEFFICIENT VALUES

The Mobility G_i—We shall now attempt to give a little deeper meaning to the mobility. The diffusion flow in solids is the observed net effect of a large number of individual movements of atoms between successive positions of equilibrium in the lattice, and it is only when there is a gradient of concentration in the alloy that these individual atomic migrations manifest themselves as a change in the concentration at points along the gradient. We require to relate G_i to some parameter or parameters of these atomic processes.

We have defined G as the average velocity of a migrating atom under unit diffusion force, *i.e.* unit free energy gradient, the average being taken over the whole time interval required for the passage of the atom from one equilibrium position to another, which includes time spent

by the atom at the first equilibrium position prior to moving to the second. If the distance between equilibrium positions is λ, then the average velocity of a diffusing atom is given by

$$v = \lambda n \qquad \ldots . \quad (36)$$

where n is the probability per second that a migration occurs along the direction of the free energy gradient. The lattice potential energy associated with any migrating atom of a homogeneous alloy in its initial (A) and final (B) equilibrium positions is the same, but midway between them the energy must rise to some maximum value E, forming a potential barrier between the two equilibrium positions as depicted in *Figure 4*. The height of the potential barrier will of course depend upon the particular process by which the migration is taking place. When a migration occurs in the absence of a concentration gradient the atom or atoms at A participating in the migration process first acquire in the course of their thermal vibrations and exchanges of energy with their neighbours an amount of energy equal to or greater than E, the activation energy for the process, and the system then proceeds over the top of the potential barrier into its new equilibrium position at B. n is quite clearly to be related to the probability p of the migration.

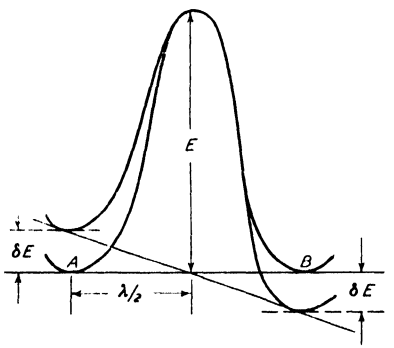


Figure 4. The effect of a concentration gradient upon the potential barrier

There are two different ways of arriving at an expression for p, as emphasized by ELEY.[49] If we assume that an equilibrium exists between atoms in their normal positions and atoms which have acquired the necessary activation energy and are in the activated state situated at the top of the barrier, and that the rate of passage of atoms from A to B is determined entirely by the rate at which atoms in this activated state pass along the top of the barrier, then we obtain the reaction rate theory, or transition state theory value of p, given by[50,51,19]

$$p = P\frac{kT}{h}\,e^{-\Delta F/RT} = P\frac{kT}{h}\cdot\frac{F^*}{F}\,e^{-E/RT} = P\frac{kT}{h}\,e^{\Delta S/R}\,e^{-\Delta H/RT}$$

$$= P\,e\frac{kT}{h}\,e^{\Delta S/R}\,e^{-E/RT} \qquad \ldots . \quad (37)$$

where ΔF, ΔS, and ΔH are the differences in free energy, entropy, and enthalpy between the equilibrium and activated states and F and F^* are the values of the partition functions for those states respectively. P is the probability per second that a new equilibrium position is available for the migrating atom to move into.

If, on the other hand, we assume that the rate of crossing the top of the barrier is very much more rapid than the rate at which the activation energy is accumulated, so that this latter is the rate controlling process, then p is proportional to the probability per unit time that the atom or atoms participating in the process possess energy greater than or equal to E, and is given by

$$p = P\nu \frac{1}{(s-1)!}\left(\frac{E}{kT}\right)^{s-1} \cdot e^{-E/kT} \quad \ldots . \quad (38)$$

where ν is the average frequency of the thermal vibrations of the atoms and s is the number of degrees of freedom over which E is accumulated. We shall refer to equation 38 as the kinetic theory value for p, after BARRER.[53]

This reduces when $s = 1$ to the more familiar

$$p = P\nu\, e^{-E/kT}$$

The essential point is, as we shall now see, that in each case p involves the height of the potential barrier in an exponential term.

Suppose now that a concentration gradient exists in the alloy at the point occupied by the atom considered. There will then also be a free energy gradient or force per atom given by:

$$\frac{1}{\mathcal{N}}\frac{\partial \mu_i}{\partial x} = kT\frac{\partial \mathcal{N}_i}{\partial x}\frac{1}{\mathcal{N}_i} \quad \text{in ideal solution}$$

or, in non-ideal solution:

$$\frac{1}{\mathcal{N}}\frac{\partial \mu_i}{\partial x} = kT\frac{\partial \mathcal{N}_i}{\partial x}\left(\frac{1}{\mathcal{N}_i} + \frac{\partial \log \gamma_i}{\partial \mathcal{N}_i}\right)$$

where $\mathcal{N}$ is Avogadro's number and $\mathcal{N}_i$ the mol fraction of i, and the effect of this will be to diminish the height of the potential barrier on the side of higher concentration and to increase it on the other side by amounts each equal to δE (see *Figure 4*), where

$$\delta E = \frac{1}{\mathcal{N}} \cdot \frac{\partial \mu}{\partial x} \cdot \frac{\lambda}{2}$$

The probability per second of migrations from left to right in the diagram now exceeds the corresponding probability of migrations from

right to left so that there is a net flow of atoms down the free energy gradient, which of course constitutes the observed diffusion flow, and the number of migrations per second is given by:

$$\begin{aligned} \mathrm{n} &= \frac{1}{m} A \left(\mathrm{e}^{\frac{-(E-\delta E)}{kT}} - \mathrm{e}^{\frac{-(E+\delta E)}{kT}} \right) \\ &= \frac{1}{m} A 2 \frac{\delta E}{kT} \cdot \mathrm{e}^{-E/kT} \quad \text{since } \delta E << kT \\ &= \frac{1}{kT} \frac{A}{m} \mathrm{e}^{-E/kT} \cdot \frac{1}{\mathcal{N}} \frac{\partial \mu}{\partial x} \lambda \qquad \ldots . \quad (39) \end{aligned}$$

where A is to be taken from any of the above expressions for p. $1/m$ is the fraction of the available directions of migrations for each atom which are effectively along the concentration gradient. (In the simplest case of six equivalent directions, $m = 6$.)

Inserting the value of n (equation 39) in equation 36, we find:

$$v = \frac{1}{kT} \frac{A}{m} \lambda^2 \mathrm{e}^{-E/kT} \cdot \frac{1}{\mathcal{N}} \frac{\partial \mu}{\partial x}$$

that is, v is directly proportional to the force acting, as tacitly assumed in deriving equations 13 and 19, so that

$$\begin{aligned} G &= \frac{1}{kT} \frac{A}{m} \lambda^2 \mathrm{e}^{-E/kT} \\ \text{or} \quad GkT &= \frac{A}{m} \lambda^2 \mathrm{e}^{-E/kT} \\ \textit{i.e.} \quad GkT &= \frac{1}{m} \lambda^2 p \qquad \ldots . \quad (40) \end{aligned}$$

In the more general case, if the probability of a migratory displacement r_j is given by $\tau(r_j)$ then λ^2/m is given by

$$\lambda^2/m = \tfrac{1}{2}\langle x^2 \rangle$$

where $\langle x^2 \rangle$ is the second moment of τ (see CHANDRASEKHAR[52]).

Then $$GkT = \tfrac{1}{2}\langle x^2 \rangle p.$$

We may note that GkT is:

1. Equal to the average velocity with which atoms migrate under unit diffusional force, *i.e.* under unit free energy gradient,
2. Directly proportional to the probability of migration of an atom from its lattice position, even in the absence of a concentration gradient. For this reason GkT is often called the place change number (*Platzwechselzahl*), and

3 Equal to the measured diffusion coefficient when the solution may be assumed ideal.

Inserting into equation 40 the value of p according to the assumptions made about the crossing of the potential barrier, and the expression for P appropriate to the particular mechanism of diffusion we are considering, we can obtain from equation 40 most of the expressions usually given for the ideal solution value of the diffusion coefficient, that is for GkT.

Atomic Interchange Mechanisms—When atoms migrate through the lattice by exchanging their positions with atoms on neighbouring lattice sites, the value of P will be unity, and the expression for the ideal solution value of the diffusion coefficient becomes, on the kinetic theory (see BARRER[53]):

$$GkT = D_{id} = \tfrac{1}{6}\lambda^2\nu \frac{1}{(s-1)!}\left(\frac{E}{RT}\right)^{s-1} . e^{-E/RT} \quad . . . \quad (41)$$

where E is now expressed in calories per gram atom and m is taken equal to 6. We might expect that the smallest value of s for this mechanism would be 2, one degree of freedom per atom, when equation 41 becomes

$$D_{id} = \tfrac{1}{6}\lambda^2\nu\left(\frac{E}{RT}\right) e^{-E/RT}$$

an expression given by BRADLEY.[54]

Equation 41, however, ignores the possibility that before an atomic exchange mechanism can operate a synchronous outward movement of atoms bordering on the exchanging pair might be necessary. This has been discussed by CICHOKI[55] and by BARRER.[51,53] If the probability of one atom moving in the correct direction is $1/a$, and if n atoms must move simultaneously, the overall probability of the synchronized outward movement of these n atoms is $(1/a)^n.p$ and therefore D must now be multiplied by this to give

$$D_{id} = \left(\frac{1}{a}\right)^n \tfrac{1}{6}\lambda^2\nu \frac{1}{(s-1)!}\left(\frac{E}{RT}\right)^{s-1} . e^{-E/RT} \quad . . . . \quad (42)$$

$$= D_0 e^{-E/RT} \quad . . . . \quad (43)$$

The transition state theory value for D_{id} when $P = 1$ is given by: [19,50]

$$D_{id} = \tfrac{1}{6}\lambda^2 \frac{kT}{h} e^{\Delta S/R} . e^{-\Delta H/RT} = \tfrac{1}{6}\lambda^2 e \frac{kT}{h} e^{\Delta S/R} . e^{-E/RT} \quad . . (44)$$

$$= D_0 e^{-E/RT} \quad . . (43)$$

All experimentally determined diffusion coefficients obey a law of this form (equation 43) in which D_0 is relatively independent of T and ranges in value for different systems from about 10^{-9} to 10^4.

When the energy E is distributed over many degrees of freedom so that there is a kind of local melting around the exchanging pair, ΔS in equation 44 will be large, the term $(E/RT)^{s-1}\{1/(s-1)!\}$, which may be as high as 10^6 for suitable values of s and T, will predominate in equation 42 and a high value of D_0 will result. With $\lambda^2 \simeq 10^{-15}$ and $\nu \simeq 10^{-13}$ we find that D_0 may reach $\sim 10^4$. On the other hand, if an essential feature of the exchange process is the synchronization of outward movements of atoms bordering on the exchanging pair, the $(1/a)^n$ term in equation 42 will predominate and a small value of D_0 will result. Such a synchronization will mean an increase in order in the vibrations of the atoms concerned and so will be accompanied by a decrease in entropy. ΔS in equation 44 will then be small or even negative, and again we shall have a low value of D_0. For a face-centred cubic crystal n may be about 12 with $a = 6$, so that $(1/a)^n \sim 10^{-9}$. Thus the observed range of values of D_0 can be well accounted for on the supposition of an exchange process.

Stearn and Eyring[50] have actually calculated values of ΔS from equation 44 and compared these with the value ΔS_0 to be expected on the supposition of a simple process involving only the movements of the exchanging pair. When the calculated value of ΔS is of the same order as ΔS_0, as for the diffusion of copper, palladium, and platinum into gold, it is assumed that no great disturbance of the lattice is occurring during the diffusion process, but when ΔS is very much greater, as for example for self diffusing systems, or very much less than ΔS_0, as for the diffusion of many metals into silver, then it is assumed respectively that a local melting is occurring, or that phase relationships between atoms bordering on the neighbouring pair are being satisfied.

As Barrer points out,[51] these calculations are only significant when it is known for certain that diffusion is by a process of atomic exchange. In other cases the calculation of S would involve a knowledge of A' or A'' (see below), which at present can only be estimated to an order of magnitude.

'Hole' or 'Vacancy' Mechanisms—There exist in the lattice a very small number of holes or unoccupied lattice sites, the Schottky defects,[56]

which can move through the lattice by exchanging their positions with adjacent atoms. Whenever a hole migrates, an atom migrates in the opposite direction, and it is this displacement of atoms accompanying the diffusional motion of the holes which is responsible for the observed diffusion. Now any particular atom in the lattice can only migrate when a hole has diffused into one of the lattice sites adjacent to it, so that P is the probability that an atom has a hole as neighbour:

$$P = \mathcal{N}'/\mathcal{N}$$

where $\mathcal{N}$ = total number of atoms and $\mathcal{N}'$ = the very much smaller number of holes.

The kinetic theory value for D is then:[53]

$$D = \tfrac{1}{6}\lambda^2\nu \frac{\mathcal{N}'}{\mathcal{N}} \frac{1}{(s-1)!}\left(\frac{E'}{RT}\right)^{s-1} . e^{-E'/RT} \quad . . . . \quad (45)$$

where E' is now the activation energy which an atom must acquire before it can move into an adjacent hole. We should not expect this mechanism to produce much of a disturbance among the immediately neighbouring atoms, so that we can justifiably make the simplifying assumption that $s = 1$. Equation 45 then becomes:

$$D = \tfrac{1}{6}\lambda^2\nu \frac{\mathcal{N}'}{\mathcal{N}} e^{-E'/RT} \quad . . . . \quad (46)$$

The value of D according to the reaction rate theory is:[51]

$$D = \tfrac{1}{6}\lambda^2 \frac{kT}{h} \frac{F^*}{F} \frac{\mathcal{N}'}{\mathcal{N}} e^{-E'/RT} \quad . . . . \quad (47)$$

choosing that value of p from equations 37 most suitable for the present discussion. If we assume that in the activated or transition state a migrating atom differs from an atom in the normal state only in that it has one less vibrational degree of freedom, then:[19,50]

$$\frac{F^*}{F} = (1 - e^{-h\nu/kT})$$

and equation 47 becomes:

$$D = \tfrac{1}{6}\lambda^2 \frac{\mathcal{N}'}{\mathcal{N}} \frac{kT}{h} (1 - e^{-h\nu/kT}) e^{-E'/RT}$$

$$= \tfrac{1}{6}\lambda^2\nu \frac{\mathcal{N}'}{\mathcal{N}} e^{-E'/RT} \quad . . . . \quad (46)$$

when T is sufficiently large for kT to be much greater than $h\nu$. In their simplest forms, then, the transition state theory and kinetic theory values of D for a hole mechanism diffusion process are the same.

Now $\mathcal{N}'/\mathcal{N}$ can be calculated from thermodynamic theory (see, for example MOTT and GURNEY[57] and BARRER[51]) and when the expression for it is inserted into equation 46 this becomes:

$$D = \tfrac{1}{6}\lambda^2 \nu A' \, e^{-(U'+E')/RT} \quad \ldots\ldots (48)$$

$$= D_0 \, e^{-Q/RT} \quad \ldots\ldots (43)$$

where U' is the change in energy of the lattice following the formation of one hole and A' is a constant which takes into account the dependence of U' and E' upon temperature, pressure, and the difference in vibration frequencies of atoms which have and have not a hole as neighbour. We note that the experimentally observed activation energy Q is no longer just the activation energy required for an atom to migrate to a neighbouring equilibrium position as it was for the exchange process, but is now the sum of this and the energy of formation of a hole. Similar remarks will apply to the observed activation energy for diffusion by interstitial mechanisms.

The value of A' may vary considerably from metal to metal and so account for the observed range of values of D_0. The order of magnitude of D_0, calculated using $\nu \simeq 10^{13}$, $\lambda^2 \simeq 10^{-15}$, and Mott and Gurney's very rough estimate of A' between 10^3 and 10^4, is D_0 between 1 and 10. This is well within the range of values observed for D_0, and, perhaps fortuitously on account of the very approximate value of A', quite close to the values of D_0 found from self diffusion experiments.

The mechanism we have described, assuming it does a uniform distribution of holes and that the successive moves of any particular atom are independent of one another, is only valid strictly for such cases as self diffusion, where these assumptions may be expected to hold. We shall find later after considering some of the experimental results on diffusion in very dilute binary alloys that our model of diffusion, by this simple hole mechanism, requires considerable modification if we wish to retain the concept of holes as being primarily responsible for the observed diffusion in dilute alloys. More detailed types of vacancy mechanisms, too, will require to take into account the dependence of the concentration of vacancies upon the alloy composition, and the rate of formation and disappearance of vacancies at grain boundaries and at the metal surfaces. The papers by Fisher, Holloman, and Turnbull[34] and by Seitz[128] represent steps in this direction.

Diffusion through Interstitial Space—The third mechanism by which it is often thought diffusion in metals might occur is by the agency of atoms in interstitial positions,[58,59] and as an example of this we will consider diffusion by Frenkel defects. An atom may vacate its normal site in the lattice, so producing a hole, and take up an interstitial position. The formation of this defect (hole plus interstitial atom) will require an amount of energy U'' and is only regarded as complete when hole and interstitial atom are sufficiently far apart for there to be no reaction between them. Diffusion then occurs by the migration of interstitial atoms from one interstitial position to another, a process requiring an activation energy E'', say. Although U'' may be quite high, so that only very few interstitial atoms are formed, it may be quite easy for atoms to migrate once they have become interstitial (*i.e.* E'' may be low) on account of the lattice distortion which the introduction of an atom into an interstitial position will produce. P, the probability that any particular atom has a new equilibrium position into which it may migrate, is now the fraction of time it spends in interstitial positions or just the probability that the atom is interstitial, and so is given by $P = \mathcal{N}''/\mathcal{N}$, where $\mathcal{N}''$ is the number of interstitial atoms and $\mathcal{N}$ the total number of atoms. ($\mathcal{N}'' << \mathcal{N}$.) The kinetic theory and transition state theory values for D are then respectively:[51,53]

$$D = \tfrac{1}{6}\lambda^2\nu'' \left(\frac{E}{RT}\right)^{s-1} \cdot \frac{1}{(s-1)!} e^{-E''/RT} \frac{\mathcal{N}''}{\mathcal{N}} \quad . \quad . \quad . \quad . \quad (49)$$

$$D = \tfrac{1}{6}\lambda^2 \frac{kT}{h} \frac{F^*}{F} e^{-E/RT} \cdot \frac{\mathcal{N}''}{\mathcal{N}} \quad . \quad . \quad . \quad . \quad (50)$$

If we make the same simplifying assumptions as for the hole mechanism, then these both reduce to

$$D = \tfrac{1}{6}\lambda^2\nu'' \frac{\mathcal{N}''}{\mathcal{N}} e^{-E''/RT} \quad . \quad . \quad . \quad . \quad (51)$$

ν'' is now, of course, the vibration frequency of an interstitial atom. As before, $\mathcal{N}''/\mathcal{N}$ is readily calculable,[51,57] and when the value for it is substituted into equation 51 this becomes:

$$D = \tfrac{1}{6}\lambda^2\nu'' \sqrt{\left(\frac{\mathcal{N}_0''}{\mathcal{N}_0}\right)} A'' e^{-(\frac{1}{2}U''+E'')/RT}$$

$$= D_0 e^{-Q/RT} \quad . \quad . \quad . \quad . \quad (52)$$

where N_0'' is the total number of interstitial positions and A'' has a meaning analogous to that of A'. (BARRER[51] seems to assume that $N_0'' = N$.) MOTT and GURNEY[57] have estimated $A''\sqrt{(N_0''/N)}$ at between about 10 and 10^4, so that the order of magnitude of D_0 turns out to be from approximately 10^{-1} to 10^2 cm^2/sec.

The above process applies, of course, only to diffusion in substitutional solid solution. With interstitial solid solution, such as of carbon in iron where the solute atoms are already held in the interstices of the solvent lattice, diffusion of the solute occurs merely by migration from one interstitial position to another. In very dilute solutions there is always another vacant site for the diffusing solute atom to move into, so that $P = 1$ and the expressions for the diffusion coefficient take just the same form as for the exchange process of diffusion. For example:

$$D = \tfrac{1}{6}\lambda^2\nu\left(\frac{E}{RT}\right)^{s-1} \cdot \frac{1}{(s-1)!}\,e^{-E/RT} = D_0\,e^{-E/RT}$$

where E is now the activation energy for the passage of a solute atom from one interstitial position to another. Assuming as the simplest case $s = 1$, $\lambda^2 \simeq 10^{-15}$ and $\nu = 10^{13}$, we find $D \sim 10^{-3}$ which is of the same order as that found experimentally for diffusion in interstitial solutions.[53]

We may also look upon this process as a hole mechanism, the holes being unoccupied interstitial positions in the solvent lattice, but this is not usual. This type of interstitial or hole mechanism is perhaps, for geometrical reasons, the only conceivable one for the diffusion of the solute in interstitial alloys, so that in such systems the diffusion process is most likely of type 1. Unfortunately such systems are very few in number compared to the substitutional alloy systems.

Other Diffusion Mechanisms—A number of other suggestions have been made from time to time about the mechanism of diffusion through crystal lattices. SMEKAL,[60] for example, considers that mosaics play a leading role in diffusion processes and has described a number of mechanisms involving them which would lead to an observed diffusion, but it is perhaps a little premature to go into the details of such mechanisms when the exact nature of mosaics is still very uncertain. Any imperfections in the lattice might conceivably play a great part in determining diffusion processes by providing particularly favourable paths down which a transfer of matter can take place. SMITH[61]

considers that fissure-like imperfections are produced in the lattice by distortion due to a concentration gradient and that when one of the alloy components has an appreciable vapour pressure (*e.g.* zinc in brass) it will diffuse as a vapour down the imperfections and so give a greatly enhanced diffusion rate for that component. He considers that this might account for the observed Kirkendall effect in α brass[26] though it would not explain the same effect reported in aluminium-bronzes.[27]

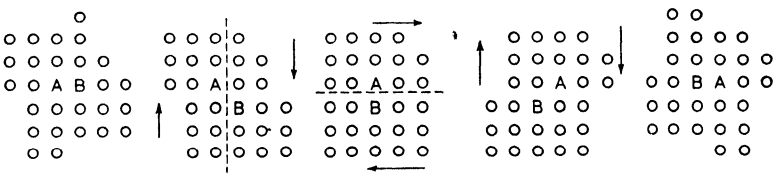


Figure 5. Bernal's model for diffusion

A number of authors have considered the possibility of movements of whole sections of the lattice with respect to one another in such a succession of ways that there results a net movement of one or more atoms through the lattice. For example, LANGMUIR as quoted by ROSENHAIN[62] suggested that diffusion occurs by the simultaneous movement in different directions of several rows of atoms. A similar process has been suggested more recently by BERNAL[63] involving spontaneous gliding of parts of the crystal over one another, which when repeated in different directions leads to an exchange of places between two atoms A and B (*Figure 5*). Such mechanisms as these seem unlikely on statistical grounds, involving as they do very extensive cooperation between large numbers of atoms.

We have now described a large number of mechanisms by which diffusion of metals in metals is thought to occur and have derived expressions for the mobility or ideal solution value of the diffusion coefficient for those processes which lend themselves to quantitative discussion, all of which are of the same form as that found for the observed temperature dependence of the measured diffusion coefficient, *viz.*

$$D_{obs} = D_0 \, e^{-Q/RT} \qquad \ldots . \quad (43)$$

In view of what has been said already, we should expect not D_{obs} but the quantity GkT, which as we have seen can be calculated from D_{obs} to follow a temperature relation of the form given in equation 43. The fact that $\log D_{obs}$ when plotted against $1/T$ always produces a

straight line, whereas it is quite clear on theoretical grounds that only $\log \{D/(1 + \partial \log \gamma/\partial \log N)\}$ should do so, is due partly to the small temperature variation in $\{1 + (\partial \log \gamma/\partial \log N)\}$ or to it being negligibly small for measurements of D at low concentration for example, and partly to the experimental measurements not being sufficiently precise to bring out the departure from linearity in the former case.

In each quantitative treatment of a diffusion mechanism we have derived two alternative expressions for D, namely the transition state theory and the kinetic theory values corresponding to the alternative assumptions made about the mechanism of crossing the potential barrier. ELEY[49] has shown that we may decide between these two alternatives by measuring $\partial E/\partial T$, the temperature coefficient of the activation energy for the migration process, since the transition state theory predicts that $\partial E/\partial T$ will be positive and the kinetic theory that it will be negative. However, such a test could only be applied to measurements made on a system in which it is known for certain that $P = 1$, for example an atomic exchange process, for only then is the observed activation energy equal to E. Apart from this limitation, the present accuracy of measurements of E hardly allow its temperature dependence to be calculated.

We have not by any means exhausted all the possible mechanisms for diffusion. We may have to consider other simple models and may find also that the simple models we have described so far need considerable modification before they represent faithfully the actual mechanisms taking place during diffusion. One such modification to the hole mechanism has already been made, and we shall discuss this more carefully when we have considered the experimental results which have made it necessary.

The theoretical non-ideal value for diffusion by any process is given merely by substituting the appropriate value for GkT into equation 13 or 19. For example, for diffusion by a hole mechanism we obtain:

$$D = \tfrac{1}{6}\lambda^2 \frac{kT}{h}\frac{F^*}{F} A' \, e^{(E'+U')/RT} \cdot \left(1 + \frac{\partial \log \gamma}{\partial \log N}\right) \quad . \; . \; . \quad (53)$$

We can see that no such simple expression as $GkT = D_0\gamma_A\gamma_B$ will be expected to account for the variation with concentration of the mobility term. Such a variation is to be looked for in the dependence of E (or ΔF, ΔS *etc*), U', A', F, F^* *etc* and to a less extent of λ^2, upon

the concentration. We also see that there is no reason for supposing that in a type 1 mechanism the mobilities of A and B atoms should be the same, if only from the fact that we should not expect the activation energy for migration of an A atom to be the same as that for a B atom migrating in the same lattice, except perhaps where the size and outer-electron configurations of A and B are very nearly the same.

One of the main problems of the theory of diffusion is to discover the nature of the diffusion mechanism in any particular case. Now we have seen that phenomenological considerations may assist in deciding whether a mechanism is of type 1 or 2, but for a knowledge of the actual mechanism we need to examine further the mobility term for the various processes. It would help if we could evaluate more accurately the various terms such as A'', A', ν'', ν' *etc* which occur in the theoretical expressions for D_0, but for the present we are faced with the difficulty that at least over a considerable range of values of D_0 any particular observed value of D_0 can be obtained on the supposition of any mechanism, especially if we make further assumptions about the conditions under which the process occurs, *e.g.* the need to satisfy phase relationships between vibrations of neighbouring atoms, the value of s, and so on.

Thus the exchange process can provide any observed value of D_0, while the interstitial and hole mechanisms give values of D_0 between 10^{-1} and 10^{2}. In view of the very rough estimates of A' and A'' and the assumption that in the latter mechanisms the process is uninfluenced by the behaviour of neighbouring atoms, we could assume for them a much wider range of values. But we have in Q, the energy of activation, a quantity which is far more characteristic of the process to which it refers, and it is by estimating theoretically values of Q for different models and comparing them with the observed value that reliable information on the nature of the mechanism may be derived. This has actually been done for self diffusion in copper. HUNTINGTON and SEITZ[64] have calculated, from the standpoint of the modern theory of metals, values of Q for the three mechanisms we have described and compared their results with the experimental value of Q obtained by STEIGMAN, SHOCKLEY, and NIX.[65] HUNTINGTON[66] has also calculated in a more detailed manner the hole mechanism value for Q. We see from *Table II* that the hole mechanism is much preferred and that the calculated and observed values of Q agree fairly well. Huntington and

Seitz also discuss qualitatively the possible diffusion mechanisms in a few metals near to copper in the periodic table. For the case of zinc they find that a hole mechanism would account for the observed anisotropy of self diffusion in that metal[9] but that for the alkali metals the interstitial and hole mechanisms would be in much closer competition than in copper.

Table II. Activation Energies for Self Diffusion of Copper

Q (*observed*) *Electron Volt*	Q (*calculated*) *Electron Volt*		
	Direct Exchange	*Interstitial*	*Hole Mechanism*
$2{\cdot}6^{65}$	>8 —	$\simeq 9$ —	$1{\cdot}0^{64}$ $<3{\cdot}52^{66}$

In performing similar calculations for Q for the interdiffusion of two metals one would have to take into account, among other things, any modifications to the simple models which experimental results for these systems may demand. No such calculations have yet been made.

Qualitatively it is often argued that the interstitial mechanism would be rather unlikely on account of the very high value of Q. In metals and alloys where the atom sizes are approximately the same there would be no room for the atom to take up an interstitial position without a considerable distortion of the lattice. The exchange process has also been criticized for similar reasons, for the interchange of position of two atoms, involving their squeezing past each other, will require a considerable activation energy and so be relatively unlikely. These criticisms would seem to be borne out by the above calculations for copper. The general opinion is that in highly coordinated alloy structures where the atom sizes are approximately equal diffusion takes place by a hole mechanism which involves the minimum of disturbance to the lattice, although conclusions based on such simple conceptions as these cannot be regarded as final. The confirmation of the Kirkendall effect in any large number of alloys will indicate the prevalence of a type 1 mechanism, and of these, for the reasons just mentioned, a vacancy mechanism is preferred. There is certainly a great need for any experiments, such as for example of the Johnson type, which will help to settle this question.

EXPERIMENTAL PROCEDURES AND TECHNIQUES

Procedures—We have seen that if we assume negligible changes in molal volume over the concentration range studied diffusion phenomena can be represented by an equation of the form

$$\frac{\partial c_i}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial c_i}{\partial x}\right) \quad \ldots . \quad (54)$$

where for binary systems, and for ternary systems in which type 1 mechanism operates, $D = GkT\{1 + (\partial \log \gamma_i/\partial \log N_i)\}.D$, the diffusion coefficient, is a function of temperature and composition. In the following treatment we shall restrict ourselves to binary systems. We must be prepared to find that in the general case two diffusion coefficients D_A and D_B are required, one for each component, in order to describe fully the diffusion process. We must then have available a method of solving two equations of the form 54 in such a way as to yield the two required values of D. DARKEN[37] has recently described such a method.

The usual experimental procedure for determining the diffusion coefficient of metals in metals and their variation with concentration is to place the two metals or alloys, usually in the form of thin flat cylinders, into intimate contact, to heat at some fixed temperatures for a definite time, and then to determine by some means the concentration of the two metals at points throughout the couple. When there is a complete series of solid solution between the two metals, the cylinders A and B can each be of the pure metal, but in other cases A may be of pure metal and B preferably of an alloy containing less of the second metal than the limit of its solubility in the first. This avoids the complication of phase boundaries in the diffusion layer. Alternatively, A and B may each be alloys of compositions within the limits of any single phase. Let us assume that the metal cylinders are sufficiently thick for there to be no change in concentration at their outer ends during the time for which diffusion is being followed and consider a set of axes with the origin fixed in one of the end faces and the x axis parallel to the direction along which diffusion is taking place. When the diffusion coefficients of the two components are not equal, then at any point along the x direction there will be, superimposed upon the separate and oppositely directed diffusion flows of the two components relative to some lattice point, a general mass flow which is the same for all

lattice points and which will be related in some way to the difference in the two diffusion coefficients, for when the coefficients are equal there is no mass flow. Now we are only interested, from the theoretical point of view, in the rate of diffusion of atoms past lattice points, *i.e.* in the motion of atoms relative to the lattice, so that we need some means of observing the mass flow of the lattice in order that the observed diffusion may be referred to axes fixed in the lattice. We can never observe a lattice point, and the best we can do is to refer the diffusion to some large 'atoms' or inclusions, such as the fine molybdenum wires used by Smigelskas and Kirkendall, which we must assume will remain fixed relative to, and share in the motion of, the lattice while not taking any part in the diffusion process on account of their size. Since in solid diffusion very few ($\sim 10^{-6}$) of the atoms are involved at any one time in the diffusion process, this is a valid assumption.

Suppose that the velocity of an inclusion at any point x, relative to our chosen axes fixed in an end face of the diffusion couple, is v. Then the flow of atoms of A per unit time through unit area placed perpendicular and fixed relative to the x axis is:

$$(\Delta n_{\mathrm{A}})_x = -D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} + n_{\mathrm{A}}v$$

where n_{A} is, as before, the number of atoms of A per unit volume. The flow through a similar area at $x + \mathrm{d}x$ is given by:

$$(\Delta n_{\mathrm{A}})_{x+\mathrm{d}x} = -\left(D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} - n_{\mathrm{A}}v\right) + \frac{\partial}{\partial x}\left\{-\left(D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} - n_{\mathrm{A}}v\right)\right\}$$

so that the rate of accumulation of A atoms at x becomes:

$$\frac{\partial n_{\mathrm{A}}}{\partial t} = \frac{\partial}{\partial x}\left(D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} - n_{\mathrm{A}}v\right) \quad \ldots . \quad (55)$$

Similarly for B atoms:

$$\frac{\partial n_{\mathrm{B}}}{\partial t} = \frac{\partial}{\partial x}\left(D_{\mathrm{B}}\frac{\partial n_{\mathrm{B}}}{\partial x} - n_{\mathrm{B}}v\right) \quad \ldots . \quad (56)$$

Adding equations 55 and 56 and assuming that $n_{\mathrm{A}} + n_{\mathrm{B}} = \text{constant}$, we obtain:

$$\frac{\partial}{\partial x}\left\{D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} + D_{\mathrm{B}}\frac{\partial n_{\mathrm{B}}}{\partial x} - v(n_{\mathrm{A}} + n_{\mathrm{B}})\right\} = 0$$

This may be integrated with respect to x to give:

$$D_{\mathrm{A}}\frac{\partial n_{\mathrm{A}}}{\partial x} + D_{\mathrm{B}}\frac{\partial n_{\mathrm{B}}}{\partial x} - v(n_{\mathrm{A}} + n_{\mathrm{B}}) = \varphi(t) \quad \ldots . \quad (57)$$

At $x = 0$, $v = 0$ and $\partial n/\partial x = 0$ for all t, so that $\varphi(t) = 0$. Note also that $v = 0$ at $x = \infty$ if $n_A + n_B =$ constant, so that we can check the validity of this assumption by measuring the relative displacement of the two end faces of the couple.

We have now from equation 57:

$$v = \frac{1}{n_A + n_B}\left(D_A \frac{\partial n_A}{\partial x} + D_B \frac{\partial n_B}{\partial x}\right) = (D_A - D_B)\frac{\partial \mathcal{N}_A}{\partial x} \quad \ldots . \quad (58)$$

and

$$\frac{\partial \mathcal{N}_A}{\partial t} = \frac{\partial}{\partial x}\left(D_A \frac{\partial \mathcal{N}_A}{\partial x} - \mathcal{N}_A D_A \frac{\partial \mathcal{N}_A}{\partial x} + \mathcal{N}_A D_B \frac{\partial \mathcal{N}_A}{\partial x}\right)$$

$$= \frac{\partial}{\partial x}\left\{(\mathcal{N}_A D_B + \mathcal{N}_B D_A)\frac{\partial \mathcal{N}_A}{\partial x}\right\} \quad \ldots . \quad (59)$$

where $\mathcal{N}_A$ is the mol fraction of A.

Now equation 59 is of just the same form as equation 2, for which a solution has been given by Boltzmann for boundary conditions which are satisfied by the experimental arrangement we have briefly described.

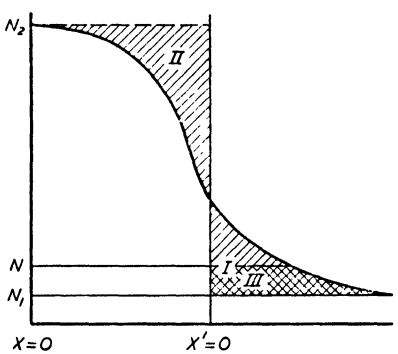


Figure 6. Concentration/distance curve

The origin is first changed from $x = 0$ to $x' = 0$, where $x' = 0$ is defined by the relation:

$$\int_{\mathcal{N}_1}^{\mathcal{N}_2} x' \mathrm{d}\mathcal{N} = 0 \quad \ldots . \quad (60)$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are the mol fractions of one of the alloy components in the two alloys or metals at time $t = 0$. After diffusion has occurred for some time t, the rough shape of the concentration/distance curve is as shown in *Figure 6*, so that relation 60 means that the ordinate $x' = 0$ is to be chosen so that the areas *I* and *II* are equal.

The plane $x' = 0$, the 'Matano Interface' is also that plane through which an equal number of atoms have passed in each direction, so that in view of the possibility of D_A and D_B not being equal we should not in general expect it to coincide with the initial position of the plane of contact of the two alloys or metals at $t = 0$. That it does not in fact do so has often been observed but no particular significance has been attached to the fact until recently, least of all was it suggested that it might be due to an inequality of D_A and D_B. If there is any appreciable

variation of the partial molal volumes of the components with composition, then this alone will produce a shift in the original interface during the diffusion, but the extent of this is usually calculable. We shall return to this point later.

We now make the assumption that N_A is a function of a variable λ, given by $\lambda = x't^{-\frac{1}{2}}$. Transforming equation 59 we obtain:

$$-\tfrac{1}{2}\lambda\partial N_A = \partial\left\{(N_A D_B + N_B D_A)\frac{\partial N_A}{\partial\lambda}\right\}$$

which on integration gives

$$-\tfrac{1}{2}\int_1^2 \lambda \mathrm{d}N = \left(\tilde{D}\frac{\partial N_A}{\partial\lambda}\right)_2 - \left(\tilde{D}\frac{\partial N_A}{\partial\lambda}\right)_1$$

where we have written

$$\tilde{D} = N_A D_B + N_B D_A \qquad \ldots\ (61)$$

Taking N_1 as the upper limit of integration, the value of $\tilde{D}$ at any other concentration N is given by:

$$\tilde{D} = -\tfrac{1}{2}\left(\frac{\partial\lambda}{\partial N}\right) N\int_N^{N_1}\lambda \mathrm{d}N$$

$$\text{or}\quad \tilde{D} = -\frac{1}{2t}\left(\frac{\partial x'}{\partial N}\right)\int_N^{N_1} x'\mathrm{d}N$$

so that a measurement of an area such as *III* in *Figure 6* and of the slope of the (N, x') curve at N suffices to determine $\tilde{D}$ at that concentration. The least accurate results are obtained for concentrations near to N_1 and N_2, that is, near the tails of the concentration/distance curve.

The validity of the assumption that $N_A = f(\lambda)$ is tested usually by plotting, for constant concentration, values of x' against $t^{\frac{1}{2}}$. For a number of different concentrations a fairly good set of straight lines through the origin is usually obtained.

We may mention also the calculation of $\tilde{D}$ at various concentrations by the method of Grube and Jedele,[13] which though only approximate because it employs the simple form of Fick's equation

$$\frac{\partial N}{\partial t} = \tilde{D}\frac{\partial^2 N}{\partial x^2} \qquad \ldots\ (3)$$

gives results in fair agreement with the more accurate Matano-Boltzmann method in those cases where the variation of D with

concentration is slight. They employ the appropriate solution of equation 3, namely:

$$\mathcal{N} = \frac{\mathcal{N}_0}{2}\left\{1 - \operatorname{erf}\left(\frac{x}{2\sqrt{\tilde{D}t}}\right)\right\} \quad \ldots . \quad (62)$$

where N_0 is the concentration of one component of the alloy at $t = 0$ and $\mathcal{N}$ the concentration at the point x after time t. The plane $x = 0$ (Grube interface) is that which intersects the $(\mathcal{N}, x)$ curve at $\mathcal{N} = \frac{1}{2}$. Values of D are then calculated by solving the equation for corresponding pairs of values of x and $\mathcal{N}$.

Other methods of solving equation 38 have been given by HOPKINS[67] and by AWBERY,[68] in which assumptions are made as to the functional dependence of D upon concentration. The solutions obtained then provide the constants in the assumed function. These methods, however, do not appear to provide any particular advantages over the one above.

Barrer[69] more recently has shown how the value of $\tilde{D}$ in equation 54 may be obtained by measurements upon the steady state diffusion of the one material through the other, that is by observing the rate of flow of A by diffusion through B, when the concentration of A and B at each point are kept constant. While valuable for the study of systems in which only one component undergoes a diffusional motion, such as for example gases and liquids through solids, the method is not applicable to the interdiffusion of two metals simply because of the difficulty of establishing a stationary flow in such systems. It has been applied by HARRIS[70] to the measurement of the diffusion coefficient of carbon through iron.

Thus from a knowledge of the concentration/distance curve after a time t we can calculate the values of $\tilde{D}$ (or $\mathcal{N}_A D_B + \mathcal{N}_B D_A$) over a range of concentration.

We may refer to the observed $\tilde{D}$ as the interdiffusion coefficient, and when written in full this becomes:

$$\tilde{D} = \mathcal{N}_A D_B + \mathcal{N}_B D_A = (\mathcal{N}_A G_B + \mathcal{N}_B G_A) kT \left(1 + \frac{\partial \log \gamma_A}{\partial \log \mathcal{N}_A}\right) \quad \ldots . \quad (63)$$

for a type 1 mechanism, and

$$\tilde{D} = (\mathcal{N}_A + \mathcal{N}_B) D_{AB} = G_{AB} kT \left(1 + \frac{\partial \log \gamma_A}{\partial \log \mathcal{N}_A}\right) \quad \ldots . \quad (64)$$

for a type 2 mechanism, since the mobilities of A and B are equal in such cases. It will be recalled that equation 63 was employed earlier in discussing Johnson's experiments.

Now the procedure which we have outlined above is that employed in all the so-called Matano-Boltzmann analyses of diffusion experiments, so that, as Darken[37] points out, all previous measurements of diffusion coefficients are correct in principle but actually only give the values of $\tilde{D}$ and its variation with concentrations. In other words, the implicit assumption has always been made in the past that diffusion phenomena in a binary system can be represented by a single diffusion coefficient, or, as we can now see, that D_A and D_B are always equal.

In the light of the above, the correct procedure to be followed in a study of diffusion in a binary system and the simplifications that can be introduced in special cases will now be outlined.

The mutual diffusion of two metals over the range of concentration covered by a single phase is studied by the method given above, thus yielding a knowledge of $\tilde{D}$ over this range of concentration. Experiments are then carried out in the way to be described later to measure the velocity v of inclusions or markers placed at various positions in the diffusion zone. When the partial molal volumes of the two components vary appreciably with composition there will be volume changes during the diffusion which alone will produce an observable motion of the markers. The extent of this displacement can be calculated when the density of the alloy over the relevant range of concentration is known, and applied as a correction to the observed v. If v is zero we can assume that D_A and D_B are equal to one another and therefore to $\tilde{D}$. In such a case we cannot say whether we are dealing with a type 1 or a type 2 mechanism, but only know that one coefficient is sufficient to describe the observed diffusion. When v is not zero, the simultaneous solution of equations 58 and 61 will yield the values of D_A and D_B over the range studied, and we can assume that we are dealing at least in part with a type 1 mechanism.

As an alternative to the above procedure for determining $\tilde{D}$ over a range of concentrations we may perform a series of experiments on diffusion between alloys of only slightly differing composition and take the value of $\tilde{D}$ obtained, assuming it to be constant over this small range of concentrations, as corresponding to the mean concentration of the two alloys.[36] We can in such cases employ the simpler form of Fick's equation:

$$\frac{\partial c}{\partial t} = \tilde{D}\frac{\partial^2 c}{\partial x^2}$$

the appropriate solution for which has already been given (equation 62):

$$c = \frac{C_0}{2}\left\{1 - \operatorname{erf}\left(\frac{x}{2\sqrt{\tilde{D}t}}\right)\right\} \qquad \ldots . \quad (62)$$

That value of $\tilde{D}$ would be chosen for which equation 62 could best be fitted to the concentration/distance curve. This method though longer has the merit of providing accurate results over the whole range of concentration studied rather than, as in the Matano-Boltzmann method, only over the centre portion of the concentration range.

It would be convenient to study the diffusion in a system consisting of a series of disks of alloys of slightly differing compositions, covering the range required, welded together in order of composition, the thicknesses of the disks being such that no change in composition would occur at their centres (Van Liempt's cascade method[124]).

A special case of this method is to use a pure metal A and a very dilute alloy of B in A. This yields a value of D''_{BA} at low concentration (if the mechanism is of type 2) or alternatively an approximate value of D'_B at zero concentration, for since $\tilde{D} = N_A D'_B + N_B D'_A$, the observed value of $\tilde{D}$ tends in the limit to the value of D'_B as N_B becomes very small and N_A becomes more nearly equal to one. We note also that since

$$D_A = GkT\left(1 + \frac{\partial \log \gamma_A}{\partial \log N_A}\right) = GkT\left(1 + N_A \frac{\partial \log \gamma_A}{\partial N_A}\right)$$

D_A will, in the limiting case of zero concentration, become equal to the mobility term $G_A kT$. These limiting values of D at low concentrations are of considerable theoretical importance,[90] so that their measurement is of some interest.

We next require a knowledge of the activities of one of the components of the alloy over the range of concentration for which D has been studied in order to calculate the value of the mobility from equation 13 or 19:

$$D_A = GkT\left(1 + \frac{\partial \log \gamma_A}{\partial \log N_A}\right)$$

We shall briefly mention later the methods available for the determination of activities and activity coefficients. As we have seen, theoretical models of the diffusion process give only the value of GkT, that is to say, they only give the value of the diffusion coefficient in ideal solutions, so that the calculation of this quantity from observed

diffusion data is of great importance. In the past, in comparing theoretical and observed values the comparison has always been of the observed diffusion coefficient with the theoretically predicted value of the mobility term GkT, which as we can now see are not the same thing. So far, fortunately, this has not mattered because the theory of diffusion is at present capable of predicting only very approximate values of the diffusion coefficient. Time, of course, will take care of this, but for the present it is most important to distinguish between the true mobility or place change number GkT and the effective diffusion coefficient D, especially in discussing diffusion rates and their dependence upon concentration, in relation to other processes in metals which involve atomic movements such as *e.g.* order-disorder transformations and slip.[71] Dehlinger[21] seems to have been the first to make this important distinction, and we have mentioned how he found that GkT and D_{obs} varied over the concentration range in quite different manners. *Figure 16* shows the value of $\tilde{D}$ as observed and after correction for the departure of the solution from ideality, plotted against concentration for the diffusion of zinc in brass.

The variation of the observed diffusion coefficient with temperature is found almost invariably to follow a law of the form:

$$D_{obs} = D_0 \, e^{-Q/RT} \qquad \ldots . \quad (43)$$

where D_0 is practically independent of temperature. In those cases where it does not do so, it can usually be shown that more than one mechanism of diffusion is operative. The value of Q, the activation energy, is found by measuring the slope of a plot of log D_{obs} against $1/T$, and is always quoted in results of experiments. Both D_0 and Q depend upon the concentration, so it is important to plot for a series of values of $1/T$, values of log D_{obs} measured all at the same concentration. Very many measurements give some mean value of D_{obs} over a range of concentration so that the activation energies calculated from these data will also be mean values over the same range.

Now theoretically calculated values of the diffusion coefficient always show a temperature dependence of the form 43, but since such calculations give only the ideal solution value GkT for the diffusion coefficient it is really only the mobility term in the observed diffusion coefficient that we should expect to depend upon temperature in the manner of equation 43. In other words, the more accurate form of

equation 43, when one coefficient suffices to describe the observed diffusion, is:

$$D_{obs} = D_0 \, e^{-Q/kT} \left(1 + \frac{\partial \log \gamma}{\partial \log N}\right) \quad . \quad . \quad . \quad . \quad (65)$$

Now Q is of some considerable theoretical importance, and since $\{1 + (\partial \log \gamma / \partial \log N)\}$ will in general vary with temperature even at constant concentration, it is important to plot, not log D_{obs}, but log $D_{obs}/\{1 + (\partial \log \gamma / \partial \log N)\}$ against $1/T$ in order to calculate at any given concentration an accurate and physically more significant value of Q. This was first pointed out by Birchenall and Mehl[18] and by Smoluchowski.[72]

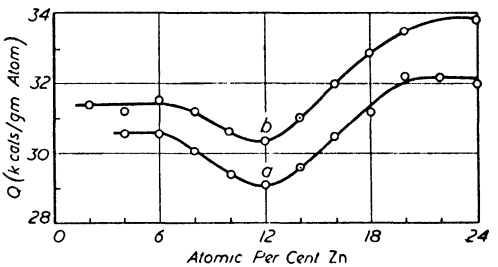


Figure 7. Activation energy Q for the diffusion of zinc in copper. Calculated a from observed diffusion coefficient and b from mobility or ideal solution value of diffusion coefficient

Figure 7 shows values of Q at different concentrations for the diffusion of zinc in α brass as calculated by the two methods, using the diffusion data of RHINES and MEHL[73] as selected by Birchenall and Mehl[18] for their paper on activities and diffusion in metallic systems, and for the second method the activity data of HARGREAVES[74] as smoothed by Fisher, Holloman, and Turnbull.[84] Following Rhines and Mehl, the lines were drawn through the most widely separated points on the temperature axis. We see that there is quite a marked difference between the two curves. Since it is very often the values of the activation energies for diffusion which are employed in detecting the operation of diffusion processes in other phenomena in metals (by comparing their activation energies with that for diffusion), it is most important that the more physically significant value of Q be used.

Equation 65 was restricted to the case where one diffusion coefficient sufficed to describe the observed diffusion. In the more general case when $D_A \neq D_B$ we shall have:

$$D_{obs} = (N_A D_B + N_B D_A) \left(1 + \frac{\partial \log \gamma}{\partial \log N}\right)$$

$$= (N_A D_{0B} e^{-Q_B/RT} + N_B D_{0A} e^{-Q_A/RT}) \left(1 + \frac{\partial \log \gamma}{\partial \log N}\right)$$

so that $\log D_{obs}/\{1 + (\partial \log \gamma/\partial \log N)\}$ plotted against $1/T$ will only provide some mean value of Q_A and Q_B. The separate values of D_A and D_B must be obtained by measurement of v before the values of Q_A and Q_B can be calculated separately. The fact that relation 43 is verified in nearly all systems investigated suggests that Q_A and Q_B cannot be very unequal. It will be apparent that Q in *Figure 7* is only some mean value of Q_{Cu} and Q_{Zn}. A series of values of D_{Cu} and D_{Zn} is not available, so to illustrate the point in question we have plotted $\log D_{obs}$ and $\log D_{obs}/\{1 + (\partial \log \gamma/\partial \log N)\}$ against $1/T$.

In self diffusion experiments we have a particularly simple case, for we are observing the interdiffusion of two different isotopes of the same substance. Since G_A and G_B will be nearly equal and independent of concentration, and since γ is everywhere unity, equation 59 reduces to:

$$\frac{\partial N}{\partial t} = D \frac{\partial^2 N}{\partial x^2} \quad \ldots . \quad (66)$$

and equations 63 and 64 to:

$$D = GkT \quad \ldots . \quad (67)$$

so that the observed diffusion coefficient is the true mobility term or place change number. These equations 66 and 67 are, of course, the usual elementary ideal solution equations in terms of which most diffusion experiments in the past have been interpreted. We see now that they are only strictly applicable in the general case to self diffusing systems. However, G_A and G_B will not be exactly equal because of the slight difference in mass between the interdiffusing isotopes, and Johnson[75] by comparing the rates of diffusion of several stable isotopes of nickel in copper has shown that the mobilities may be taken as inversely proportional to the square root of the isotope mass. This result can be used to correct measured self diffusion coefficients of the radioactive atoms to give the true self diffusion coefficient of the inert material.

Techniques—The experimental techniques required for the measurement of diffusion coefficients are fairly straightforward if a little laborious, but there are a few points which do not always receive sufficient attention, and as a result the value of some of the published work is difficult to assess.

Well annealed and perfectly homogeneous materials should, of course, be used in all diffusion experiments, and the importance of their purity has already been stressed.

In cases where grain boundary diffusion plays a part the observed diffusion coefficient will vary with grain size, and such a variation must be investigated if we are to know to what extent the measured D is due to lattice diffusion through the grains and to what extent due to diffusion down the grain boundaries; far too many workers have ignored this point altogether. When there is no grain boundary diffusion the measured D will of course be independent of grain size. Grain boundary diffusion can also be detected by metallographic examination of the diffusion zone. Only when there is no preferential penetration down grain boundaries will the lines or planes of constant composition, as revealed by careful etching, pass straight across grain boundaries (see *e.g.* Figure 6, of reference 76). Lattice diffusion can usually be made to predominate by working at sufficiently high temperatures but it is preferable to perform experiments on single crystals so as to observe the lattice diffusion only and then to relate these measurements to further ones on polycrystalline material.[77]

Intimate atomic contact between the two alloys or pieces of metal between which diffusion is to take place is achieved where possible by electroplating[13,73,75] or alternatively and where electroplating is not applicable, by welding,[36,78,82] or by rolling or drawing.[14,15,79] Mehl[4] has drawn attention to the importance of oxide films or other impediments in the plane of contact where, if these are not quickly dissolved during the diffusion anneal, the effect will be to diminish the effective area of contact and so give a reduced value of D. A similar result would occur if the rates of diffusion of one or both of the components through the oxide layer were very much less than through the bulk of the material. He also points out that the method of producing intimate contact by rolling or drawing offers special advantages in fragmenting and attenuating the oxide layers. The possible presence and effects of such oxide films are not always investigated or even discussed in connection with diffusion measurements.

The difficulty could be avoided altogether if it were possible by some means to produce a concentration gradient in an initially homogeneous alloy and then observe the process of rehomogenization by diffusion, although the interpretation of the results would be more difficult. Such a specimen might be prepared by cooling a molten alloy by very slowly drawing it out of a furnace, when the composition of the solidifying alloy will follow the solidus curve.

No experiments have yet been done on the rates of interdiffusion of two single metal crystals, but should such be carried out it might be important to specify carefully the relative orientation of the two crystals, for there is some evidence[95] that this would determine the observed rate of diffusion to a considerable extent.

We have seen that for a complete investigation on the dependence of $\bar{D}$ upon concentration a knowledge of the concentration at points throughout the diffusion zone is essential. We might add that measured concentration/distance curves are not always interpreted in the manner described to give $\bar{D}$ as a function of concentration, but only used to provide an average value of the diffusion coefficient. Successive thin layers of material parallel to the original interface may be removed on a lathe and their composition determined by any convenient method such as chemical,[15,73,113] quantitative spectrographic,[15,78,80] or x-ray analysis.[81,82] PAIĆ[83] has described a method of determining the composition by microradiography. If one of the metals or alloys contains an added amount of a radioactive isotope of one of the components, the composition of successive layers can be determined by measurement of their activity, but this method is usually reserved for self diffusion measurements. Mass spectroscopic analysis of thin layers has been employed by Johnson[75] in his measurements on the diffusion rates of the stable isotopes of nickel into copper. Alternatively, the composition of the new surfaces, revealed by the removal of the layers, may be determined by taking x-ray reflection photographs. By using very soft x-rays, or better still electrons, the composition measured is an average value over only a very shallow layer of the surface. The removal of extremely thin layers of material, as is necessary when the diffusion is slight, can be effected in some cases by electrolytic etching.[84] The composition of the new surface is then determined in the above manner. BÜCKLE[85] has developed a method of determining the concentration of alloys which depends upon the variation with composition of the microhardness of solid solutions and has employed this to obtain the concentration/depth curves for various metals diffusing into aluminium.[86] A rather ingenious micrographic method, which avoids the necessity of removing layers of material from the specimen, has been used by Brick and Phillips[10] to obtain the concentration/penetration curves for the interdiffusion of copper and of magnesium with aluminium.

There are several other methods of following diffusion rates which do not depend upon a knowledge of the concentration/depth curve but which, since it is assumed in their interpretation that D is independent of concentration, are strictly only applicable to measurements on the diffusions between a metal and a very dilute alloy of that metal with another, or between two alloys of nearly equal concentration. These limitations are by no means always imposed but the methods at least provide a rough measure of the magnitude of the diffusion coefficient for the system investigated, although the actual value obtained for D may only be a mean over some range of concentration.

Evaporation Method—The rate at which the more volatile component of a binary alloy diffuses out of the material when it is heated is measured as the loss in weight of the specimen after various periods of heating *in vacuo*. The method, first employed by DUNN[87] for studying zinc diffusion in α brass, is very popular among Russian workers and has been critically discussed by GERZRICKEN, FAINGOLD, ILKEVICH, and SAKHAROV.[88] It has to be remembered that the rate of loss of the volatile component is controlled not only by the rate of diffusion to the surface but also by the rate of evaporation from the surface, which must be measured separately. There are other complications in all but single crystals such as the shrinkage of grains due to the loss of the volatile component, followed by preferential diffusion along the grain boundaries which might not exist under other conditions.[4] The method has been used recently by GERZRICKEN and GOLUBENKO[89] to follow the effects of stress and plastic strain upon the diffusion of zinc out of brass and by GERZRICKEN and DEGHTYAR[48] to study the influence of magnesium and antimony on the diffusion of mercury out of mercury-lead alloys. VAN LIEMPT[90] has also used the method to measure the diffusion of iron in very dilute solution in tungsten.

Any surface property of a metal which depends upon the concentration at the surface may be used to follow the diffusion of material from the interior of a specimen to the surface. Thus the diffusion coefficient of thorium through crystals, along grain boundaries and over the surface of tungsten, has been measured by observing the concomitant changes in ionic emission from the surface thorium.[77,91,92] The photoemission properties of sodium were employed by BOSWORTH[93] to study the diffusion of sodium through and over the surface of tungsten. COLEMAN and YEAGLEY[94] evaporated a thin film first of one

metal and then another on to a glass plate and followed the interdiffusion of the two metals by observations on the changes in reflectivity of one surface. This method is very sensitive and makes possible an observation of diffusion in some cases at temperatures as low as 50°C and in times as short as 5 minutes, but the results need careful interpretation before they can be compared with measurements made on material in bulk[95] because of the effects of preferred orientation in the deposited metal layers and because the physical properties of thin films are not necessarily the same as those of the material in bulk. DUMOND and YOUTZ[96] and BURR, COLEMAN, and DAVEY[97] have followed, by x-ray diffraction methods, diffusion in similarly prepared specimens of metal films deposited on glass.

Finally there are methods available which consist in following some bulk property of the diffusing system, which in these cases often consists of specimens of many layers formed either by electroplating first one metal and then the other, and so on, or by rolling a stack of metal foils. Thus for example TANAKA and MATANO[98] studied the interdiffusion of copper and gold by measurements of the change in electrical resistance of such specimens, and MATANO,[99] by following changes in the x-ray diffraction pattern, measured the diffusion coefficient of nickel in copper. CHEVENARD and WACHÉ[100,101] and CHEVENARD and PORTEVIN[102] have described a thermomagnetic method applicable only to ferromagnetic solid solutions, in which the concentration changes are followed as changes in the Curie point. Such methods as these are convenient in that they allow the diffusion to be followed without destruction of the specimen so that many observations may be taken on the same specimen.

These methods all give values of $\tilde{D}$ (or a mean value of $\tilde{D}$ over some range of concentration). In the general case, in order to calculate from $\tilde{D}$ the separate values of D_1 and D_2 when these are not equal, we need to know the velocity of and the concentration gradient at the position of markers placed at various positions along the direction of diffusion so as to be able to solve simultaneously equations 58 and 61. The only reported measurements of v are by Darken[37] from the data of Smigelskas and Kirkendall,[26] whose experimental procedure we have already briefly described. The distance between the markers is plotted as a function of time, and from the slope of this curve is calculated the velocity of the markers at the concentration obtaining in

their immediate vicinity. Smigelskas and Kirkendall's experiments give only the value of v at one concentration, namely that at the brass/copper interface, but the method could clearly be extended to give values of v at other concentrations simply by embedding the molybdenum wires at other positions away from the interface. When this is impracticable, as for example when the metals cannot be conveniently electroplated, it will be necessary to measure the velocity of markers at the interfaces between a series of diffusion couples made up of pairs of alloys whose compositions differ only slightly and whose mean compositions cover the range for which v values are required.

If we can assume that the composition in the neighbourhood of a marker remains constant during the interdiffusion, then v may be calculated from the relation $v = (\text{shift of marker after time } t)/2t$.[37]

The assumption is readily shown to be valid when the displacement plotted against $t^{\frac{1}{2}}$ produces a straight line through the origin.[38] Observed values of v are of course to be corrected for any volume changes accompanying diffusion.

Self Diffusion—The general procedure in the measurement of self diffusion coefficients has been to cover the surface of a flat disk or plate of the metal being investigated with a radioactive isotope of the same material and then to follow the extent of the diffusion after heating by one of two methods. A concentration/distance curve is obtained, as before, by turning off in a lathe successive layers of material parallel to the original interface and measuring their content of radioactive material by means of its activity. From this D is readily calculated. See, for example, the measurements on zinc,[9] copper,[103,111] silver,[107] and gold.[104] The second method is to follow the rate of decrease of activity of the coated surface due to the disappearance of the radioactive atoms inwards by diffusion. From this, and a knowledge of the absorption coefficient of the material for the radiations emitted, D can be calculated. This method has been used in measurements on lead,[105] copper,[65,130] and gold,[106] but is generally considered to be the less accurate of the two because of the need to know these absorption coefficients. A new method involving neither the machining of numerous layers of material from the test piece nor a knowledge of any absorption coefficients has recently been described.[108] It is illustrated in *Figure 8*. The side surface of the specimen parallel to the direction of diffusion is deeply etched electrolytically to remove those layers where the

concentration of radioactive material may have been increased by surface migration, and the specimen is then placed in a lead box with the etched surface uppermost. Part of the surface is covered with a lead slide and the activity I_1 of the remaining exposed portion of the surface, of length x_1, is measured. The slide is then moved to another position, shown dotted, so that the measured activity I_2 of the now exposed portion of surface is equal to I_1. From x_1, x_2, and t, the time of the diffusion heating, D can be readily calculated.

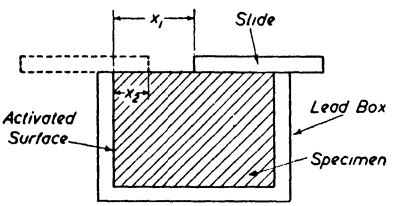


Figure 8. Apparatus for self diffusion measurements

The radioactive isotope is usually prepared beforehand and then electroplated or evaporated upon the specimen, or collected as a radiodisintegration product from a gas. Alternatively, a disk or foil of the metal, containing a proportion of the radioactive isotope, may be rolled or welded on to an inactive disk.[106,111] The method of bombarding the specimen of the metal in which diffusion is to be followed with the appropriate radiation so as to produce the radioactive isotope *in situ* has also been employed.[103,106] This has the advantage of avoiding possible effects due to surface contamination which may arise in the first method, but the disadvantage that recoil from γ or particle emission during the production of the isotope will destroy the regularity of the lattice in its immediate vicinity, and so possibly affect the observed diffusion rates.[106] Also, the initial distribution of radioactive isotope is never known with the same certainty as when it is plated as a uniform thin layer upon the surface of the specimen.

Although so far only radioactive isotopes have been used in the measurement of self diffusion coefficients, there is nothing to prevent the use of stable isotopes instead, the process of diffusion being followed by mass-spectrographic analysis of successive thin layers of material. This would be particularly useful for studying the self diffusion of magnesium for which no radioactive isotope of a sufficiently long life for a conventional self diffusion measurement is at present available.

Activity Measurements—We have seen that for a calculation of the mobility or true place change number from observed diffusion coefficients we require a knowledge of the activities of the components of the solid solution in which the diffusion is measured. WAGNER[109] has given an excellent account of the methods available for measuring

activities and activity coefficients in metallic solid solutions, and the following is a brief summary of these.

a The vapour pressure method requires a measurement of the partial vapour pressure p_A of one component A in the vapour phase over the solid solution with which it is in equilibrium. If p_A° is the vapour pressure over the pure metal A at the same temperature, and if we assume the vapour in each case to behave as an ideal gas, the activity of A is given by $a_A = p_A/p_A^\circ$ if we take the activity of the pure metal to be unity. As an example of this method we may quote Hargreaves's measurements of the activity of zinc in α brass.[74]

b The E.M.F. method involves a measurement of the E.M.F. of a cell of the type

Pure metal A	Ionic conductor of ions of metal A with Valency z	Alloy of metal B with A

in which A must be the more electropositive metal. If we transfer one gm/atom of A from right to left in the cell, the decrease in free energy is

$$\mu_A - \mu_A^\circ = RT \log a_A/a_A^\circ \quad (a_A^\circ = \text{activity of pure metal})$$

and is equal to the electrical work done $= zFE$.

$$\therefore\ E = \frac{RT}{zF} \log a_A/a_A^\circ$$

If we take the pure metal A as the standard state of unity activity, the activity of A in the alloy with B becomes

$$a_A = e^{zFE/RT}$$

There are a number of precautions to be taken, particularly in the choice of the electrolyte and the avoidance of secondary reactions at the electrodes, all of which are discussed by Wagner. Wagner and Engelhardt[39] employed this method to measure the activity of gold in alloys of gold and silver and their results were used in the interpretation of Johnson's experiments[36] on diffusion in the 50 per cent alloy.

c The chemical reaction methods consist essentially in comparing the equilibrium concentrations in a chemical reaction involving one component of the alloy with those for the same reaction in which the pure component is used. As an example we may take SMITH'S experiments,[110] from the results of which may be calculated the activity of carbon in iron. Smith measured the equilibrium partial pressures p_{CH_4} and p_{H_2} in the reactions:

$$2H_2 + C \text{ (dissolved in Fe)} \longrightarrow CH_4$$
$$2H_2 + C \text{ (graphite)} \longrightarrow CH_4$$

The equilibrium constant for the reaction is given by:

$$K = \left(\frac{p_{CH_4}}{p^2_{H_2} a_C}\right)_1 = \left(\frac{p_{CH_4}}{p^2_{H_2}}\right)_2$$

so that a_C, the activity of carbon in iron with respect to graphite, is readily calculated. A further study of the reaction of CO_2 with C dissolved in iron and with pure carbon (graphite) provided an independent set of values of a_C. These experiments have been used in the discussion of measurements of the diffusion coefficient of carbon in iron.[18,34]

EXPERIMENTAL RESULTS

The majority of experiments have provided, at best, a measure of the observed interdiffusion coefficient $\tilde{D}_{obs}$ at one or several values of the concentration, or at worst, a mean value of this $\tilde{D}_{obs}$ over some range of concentration. The difficulties in making use of many of these experimental results in any theoretical treatments of the subject are twofold. First very few of the results have been corrected for the departure of the actual solid solution from ideal behaviour, and secondly, with one exception,[26,37] no measurements have been made of v (equation 57), so that we do not know whether we are dealing with a single diffusion coefficient, that is, with a system in which the observed diffusion can be described by a single coefficient, or with a system in which two separate coefficients, one for each component, are required, so that the observed diffusion coefficient is composite ($\tilde{D} = N_A D_B + N_B D_A$). However, in the two cases of measurements of self diffusion coefficients and measurements of $\tilde{D}$ made with a low concentration of one component, *i.e.* low compared to the maximum solid solubility of that component, the results are of immediate significance, for the observed value of D actually measures the diffusion rate of one atomic species in a lattice of the same (self diffusion) or of a different species. Also, the self diffusion coefficient is the mobility or true place change number. The corrections for departure from ideality in measurements of $\tilde{D}$ in very dilute solution are usually small enough to be neglected so that again our measurements give the mobility directly.

No attempt will be made to list or even survey *all* the measurements of diffusion coefficients that have been made, for that has already been done on several occasions, as for example in the monographs by BARRER[24] or better by SEITH,[112] but a number of them will be quoted

by way of illustrating the general conclusions that have been drawn about diffusion coefficients and about the factors which are thought to influence the rates of diffusion in metals.

All measured values of $\bar{D}$ (or of $\tilde{D}$) for metals diffusing into metals, whether immediately interpretable as mobilities or whether composite in that they have not been corrected for departure from ideality or may contain two separate diffusion coefficients as type 1 mechanisms, are found to follow a temperature law of the form:

$$\bar{D} = D_0 \, e^{-Q/RT}$$

where D_0 varies for different systems from about 10^{-9} to 10^{4} and Q varies from about 10 to 115 kcal per gram atom.

The regularity with which this law is obeyed is no doubt due in large part to the experiments not being sufficiently accurate to bring out the departures from it, which the phenomenological theory we have described indicates should generally exist for measurements made on any but self diffusing or dilute solution systems.

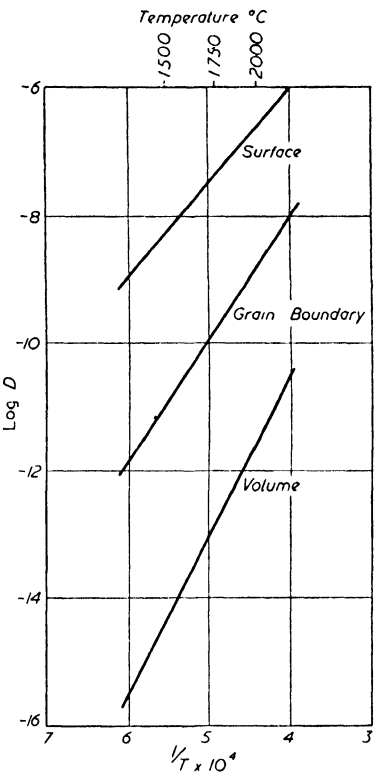


Figure 9. Diffusion of thorium in tungsten[11]

D_0 is relatively independent of temperature and is greatest for self diffusion or for diffusion between chemically and physically similar metals which may often give a continuous series of solid solution.

For the few systems that have been investigated it is found that the measured activation energy Q for lattice or volume diffusion is greater than that for grain boundary diffusion, which in turn is greater than that for surface diffusion. *Figure 9* shows log D plotted against $1/T$ for volume, grain boundary and surface diffusion of thorium in tungsten. Grain boundary diffusion is generally more important the harder the solvent, *i.e.* for such metals as tungsten or bismuth. In these cases the effect of reducing the grain size by mechanical working is to increase

considerably the observed diffusion rate, whereas for a soft metal such as lead no amount of mechanical working will affect the rates of diffusion in it.

Diffusion data have in the past often been interpreted through the semi-empirical Dushman-Langmuir equation[114]

$$D = \frac{d^2Q}{Nh} e^{-Q/RT} \quad \ldots . \quad (68)$$

$$\text{so that} \quad D_0 = \frac{d^2Q}{Nh} \quad \ldots . \quad (69)$$

where h is Planck's constant, N is Avogadro's number, and d the average interatomic spacing. Values of Q calculated from equation 68 using a single D and T value are often in very good agreement with those values obtained by plotting log D against corresponding values of $1/T$,[4,24] which is at first sight rather surprising in view of the semi-empirical derivation and simplicity of equation 68. Van Liempt[90] has, however, explained why this agreement is no proof of the validity of equation 68 and shows that the term outside the exponential in equation 68 may be changed by a factor as great as 10 without seriously affecting the value of Q calculated from it.

Less attention seems to have been given to equation 69, but Gerzricken and Deghtyar[115] have recently discussed the Dushman-Langmuir equation from this point of view and find that values of d calculated from equation 69 using observed values of D_0 and Q vary within wide limits and can be both greater and less than the lattice constant.

These results should probably not be taken as too final on account of the difficulty in computing accurate values of D_2 from experimental data, but it is extremely doubtful whether the still extensive use of equation 68 is really justified.

$\tilde{D}$ is always found to vary with concentration. Measurements over the whole concentration range of continuous solid solutions have been made on the nickel-copper system by Grube and Jedele[12,13] and by Johnson,[75] and for the alloys of gold with palladium, platinum, and nickel by Jedele.[14,116] Measurements over the α solid solution range of concentration have been made for several alloys of copper by Rhines and Mehl[73] and of aluminium by Mehl, Rhines, and Steinen[15] and by Bückle.[86] Wells and Mehl[113] have measured the diffusion rate of carbon in iron over the austenite range of concentration. Some

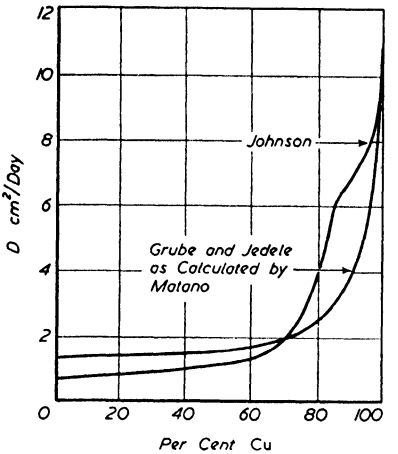


Figure 10. Diffusion of nickel and copper. Variation of $\tilde{D}$ *with concentration at 1052°C* [13,75]

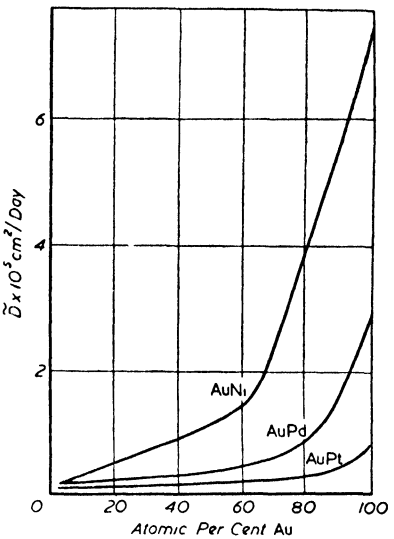


Figure 11. Diffusion coefficient $\tilde{D}$ *as a function of concentration for the systems gold-platinum, gold-palladium, and gold-nickel* [116]

of these results are shown plotted in *Figures 10* to *14*. These represent the most complete studies of diffusion constants over a range of concentration. BUGAKOV and SIROTKIN[117] have shown that for zinc and cadmium diffusing in silver the diffusion coefficient is fairly constant up to about five per cent cadmium and up to ten per cent zinc but that at higher concentrations D increases considerably. The dependence upon concentration of the diffusion rates of magnesium into lead has been demonstrated by SEITH and HERMANN.[118] They also found that the diffusion of thallium into lead was almost independent of concentration. Measurements of the rates of diffusion in β solid solution are almost non-existent, but PETRENKO and RUBINSTEIN[119] have found that for the systems silver-zinc, silver-cadmium, and copper-zinc Q is much less and diffusion much faster in the β than in the α phase region, which is as might be expected in view of the more open β structure.

When it was found that the values of D at very low concentrations for various metals diffusing in copper were approximately the same it was suggested[73] that this constant value, characteristic of the solvent copper rather than of the solutes, actually represented the self diffusion coefficient of copper. Subsequent measurement of the self diffusion of

copper appeared to vitiate this viewpoint, but Van Liempt[90] has recently reasserted that the observed diffusion coefficient of A in very dilute *substitutional* solid solution in B is equal to the self diffusion coefficient of B, and that this indicates a convenient means of measuring self diffusion coefficients. His arguments in support of this are not entirely convincing (see also reference 127), and most of the usual models for diffusion processes provide no reasons why it should be true. Although the experimental evidence would also seem to be against it,[65,118] this, as Van Liempt points out, is not altogether conclusive on account of the very large discrepancy between measurements by different observers for any one metal of its self diffusion coefficient, and of the low accuracy in the evaluation, by the Matano method, of the diffusion coefficients at low concentrations. Van Liempt also doubts whether in all systems studied for this purpose the solutions have been truly substitutional. There is clearly a need here as in the whole field of diffusion in metals for more accurate and careful measurements of diffusion coefficients.

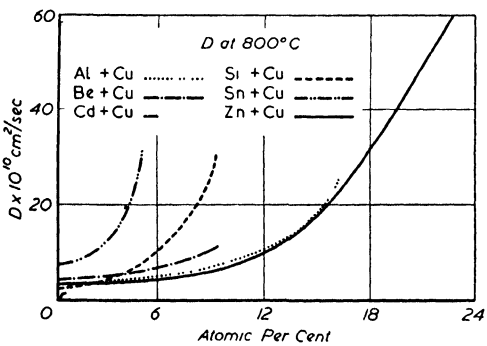


Figure 12. Selected average diffusion coefficient $\tilde{D}$, *interpolated to 800°C for systems copper-aluminium, copper-beryllium, copper-cadmium, copper-silicon, copper-tin, and copper-zinc*[73]

Rhines and Mehl[73] used their data to calculate the variation of Q with concentration in various alloys of copper. However, in view of the fact that their measurements of $\tilde{D}$ were uncorrected in any way and that their calculation of Q involves a rather arbitrary selection of data necessitated by the considerable spread of results from specimen to specimen, the particular shapes of their curves, *Figure 15*, are not to be taken as too significant. The curve for zinc-copper in *Figure 15* is quite different from that of *Figure 7* on page 353, which was drawn from calculations made on a differently selected set of data. We have already pointed out the possibly composite nature of Q in these calculations.

Bückle[86] has calculated Q at two values of the concentration for dilute solutions of copper, manganese, magnesium, and silicon in aluminium. In each case Q increases with increasing concentration.

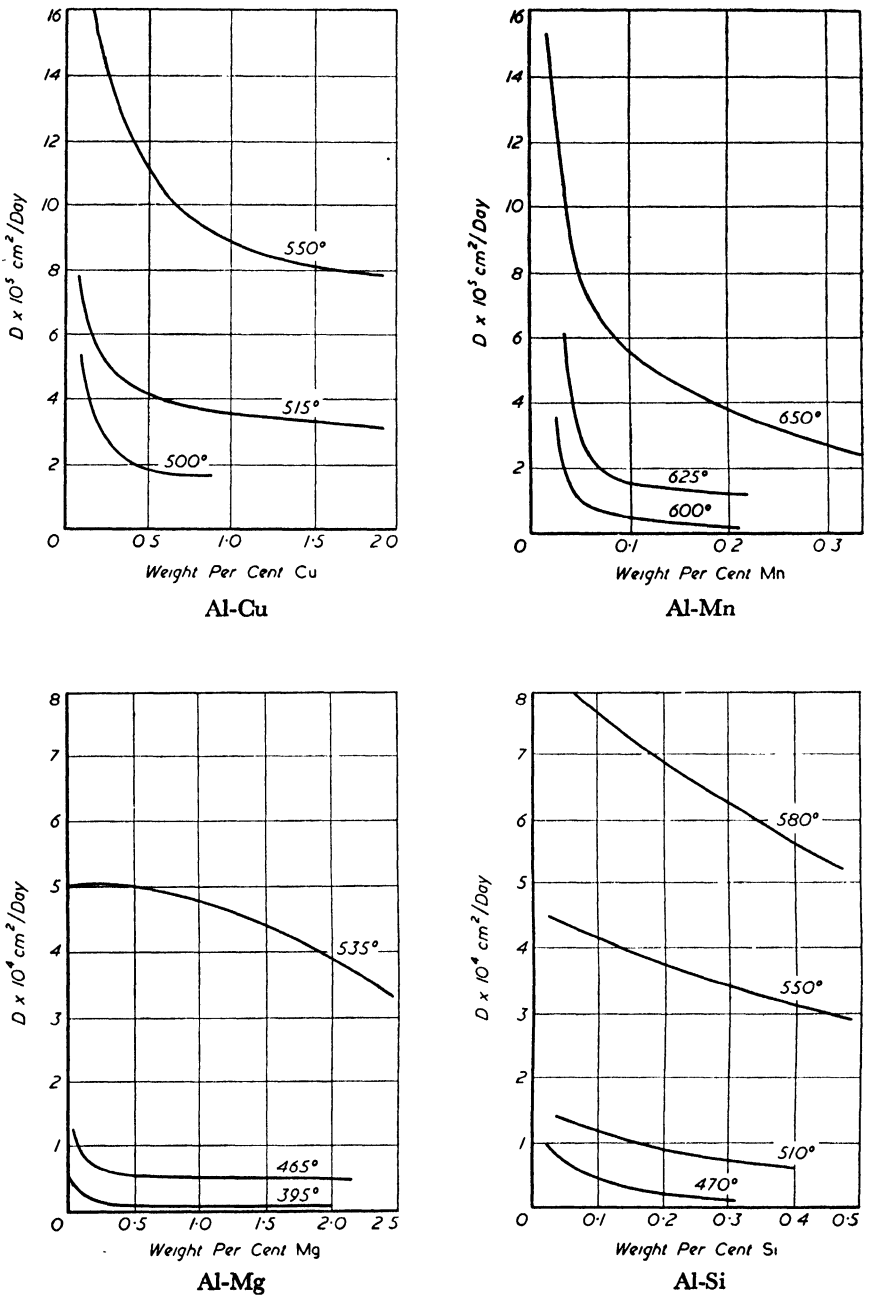


Figure 13. Diffusion coefficient $\tilde{D}$ *as a function of concentration for the alloys of aluminium with copper, manganese, magnesium, and silicon*[86]

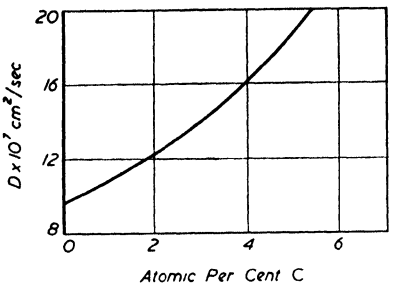


Figure 14. Diffusion of carbon in iron at 1150°C as a function of concentration [113]

There is only one reported measurement of v (equation 58), namely that of Smigelskas and Kirkendall[26] for the diffusion of zinc into copper at 785°C. Darken has calculated from their results the value of v at the concentration 22·5 per cent zinc and used this to compute the separate values of D_{Zn} and D_{Cu}. The figures Smigelskas and Kirkendall obtained were $v = 1{\cdot}24 \times 10^{-9}$ cm/sec and $dN/dx = 0{\cdot}43$ cm^{-1}.

Rhines and Mehl obtained for $\tilde{D}$ at 785°C a value of $4{\cdot}5 \times 10^{-9}$ cm^2/sec, and therefrom Darken obtained $D_{Zn} = 5{\cdot}1 \times 10^{-9}$ cm^2/sec and $D_{Cu} = 2{\cdot}2 \times 10^{-9}$ cm^2/sec. Smigelskas and Kirkendall used only one set of markers, so that only one value of v and so also only one set of values of D_{Zn} and D_{Cu} can be calculated. At very low concentrations of zinc, however, the measured diffusion coefficient is equal to the value of D_{Zn}. Rhines and Mehl give this value as $0{\cdot}3 \times 10^{-9}$ cm^2/sec, so that D_{Zn} increases seventeenfold between 0 and 22·5 per cent zinc.[37]

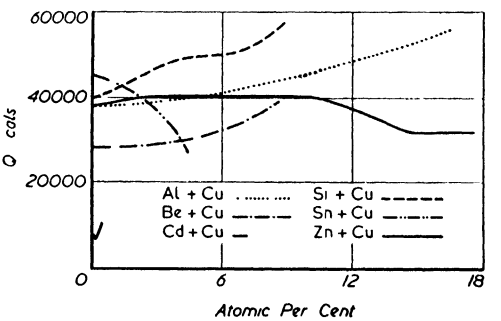


Figure 15. Relation between Q and concentration for the systems copper-aluminium, copper-beryllium, copper-cadmium, copper-silicon, copper-tin, and copper-zinc

Comparatively few measurements have been made on the activities of the components of binary metallic solid solution so that it is not possible in very many cases to correct observed values of the diffusion coefficient for departure of the solutions from ideality so as to give the mobility. The diffusion data for nickel-copper[12,13] have been very roughly corrected by Dehlinger.[21] Wells and Mehl's[113] measurements on the diffusion of carbon in iron can be corrected by employing the activity data of Smith,[110] and Hargreaves's[74] activity measurements of zinc in α brass can be used to correct Rhines and Mehl's[73] measurements of diffusion in that system. The necessary calculations have already been performed in part by Birchenall and Mehl[18] and Fisher, Holloman, and Turnbull,[34] but as we have seen these authors

made further assumptions about the mobility and their results do not provide the mobility immediately. *Figure 16* shows the observed and corrected values of D for the system zinc-copper. Hargreaves's data have been smoothed by Fisher *et al.* and their figures have been employed in the above calculations for α brass. The separate values of D_{Zn} and D_{Cu} calculated above may also be corrected to give the individual mobility values $G_{Cu}kT$ and $G_{Zn}kT$ for the zinc-copper system at a temperature of 785°C and a concentration of 22·5 per cent zinc. These are $G_{Zn}kT = 2{\cdot}57 \times 10^{-9}$ cm²/sec and $G_{Cu}kT = 1{\cdot}11 \times 10^{-9}$ cm²/sec.

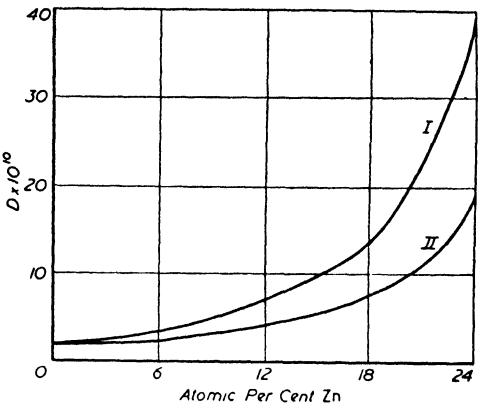


Figure 16. Diffusion data for zinc in α brass at 750°C. I observed values $= N_A D_B + N_B D_A$ *II values corrected for departure of solid solution from the ideal state* $= (N_A G_B + N_B G_A)kT$

These calculations on the zinc-copper system represent probably the most complete analysis of diffusion data in concentrated solid solution that have yet been made for any system. Much work remains to be done in extending along these lines the measurements that have already been made on many systems and in performing further such measurements on many other systems.

By studying the rates of diffusion of different metals in the same solvent attempts have been made to find correlations between the properties of the components of the diffusing system and the rates of diffusion therein.

The measurements on the diffusion of several elements in silver by SEITH and PERETTI,[78] in lead by HEVESEY and co-workers,[76] in copper by RHINES and MEHL,[73] and in gold by JEDELE[14] and by JOST[120] have been very useful in this respect, although it has not been possible to draw much more than very general conclusions from these results. Many more measurements of greater precision and more carefully interpreted than heretofore will probably be necessary before any very great progress is made towards a quantitative relationship between rates of diffusion and the properties of the components of a diffusing system.

Generally speaking, the more dissimilar the two components the faster is the rate of diffusion in any series of alloys with one metal. This dissimilarity is usually discussed in terms of degree of solid solubility, of relative position in the periodic table, and of relative size of the solute and solvent atoms, and all of these are of course linked, in some way not at all completely understood, with the differences in the electronic configurations of solute and solvent atoms and their interactions, leading to lattice distortion and polarization effects, when present together in an alloy.

Thus diffusion is usually, but not always, fastest in those systems which show the least solid solubility and accordingly slowest for self diffusion or for systems which exhibit continuous solid solubility. This is well illustrated in *Tables III* and *IV*, showing the diffusion data for lead and silver.

Table III. Diffusion of Metals in Lead[76]

Metal	Au	Ag	Cd	Bi	Tl	Sn	Pb
D cm^2/sec^{-1} (285°C)	$4{\cdot}6\times10^{-6}$	$9{\cdot}1\times10^{-8}$	2×10^{-9}	$4{\cdot}4\times10^{-10}$	$3{\cdot}1\times10^{-10}$	$1{\cdot}6\times10^{-10}$	7×10^{-11}
Q (*kcal*) ..	13	15·2	18	18·6	21·0	24·0	28·0
Crystal structure of solute	*f.c.c.*	*f.c.c.*	*c.p.h.*	*rh.*	*f.c.c.*	*tet.*	*f.c.c.*
Max. solid sol. (*atomic* %)	0·05	0·12	17	35	79	29	100
At. rad. (*Golds.*)	1·44	1·44	1·52	1·82	1·71	1·58	1·74
M.P. of solute (°C)	1,062	960	321	271	303	232	327

These two tables demonstrate the fact that roughly speaking the more removed the solute atom is from the common solvent atom in the Periodic Table, the faster is the rate of diffusion.

An increase in size of the solute atom is accompanied by an increasing rate of diffusion in alloys of silver and of copper but a decreasing rate of diffusion in lead alloys and in gold alloys (see *Table V*) but we also notice in each case that the nearer the radii of solvent and solute atoms become equal the smaller is the diffusion coefficient, although this behaviour is not quite regular.

Another interesting relationship has been observed by SEN.[121] In a

pair of metals A and B, the diffusion of A into B (at low concentration of A) is more rapid than that of B into A (at low concentration of B) when the minimum distance of approach of atoms in A is less than the same distance in B. This is illustrated by the data collected in *Table VII*, taken from Barrer[24] with additions.

Table IV. Diffusion of Metals in Silver[78]

Metal	Sb	Sn	In	Cd	Au	Pd	Ag^{107}
D cm^2/sec (760°C)	$1{\cdot}4\times10^{-9}$	$2{\cdot}3\times10^{-9}$	$1{\cdot}2\times10^{-9}$	$9{\cdot}5\times10^{-10}$	$3{\cdot}6\times10^{-10}$	$2{\cdot}4\times10^{-10}$	$1{\cdot}6\times10^{-10}$
Q (kcal) ..	21·7	21·4	24·4	22·35	26·6	20·2	45·9
Crystal structure of solute	*rh.*	*tet.*	*f.c.tet.*	*c.p.h.*	*f.c.c.*	*f.c.c.*	*f.c.c.*
Max. solid sol. (atomic %)	5	12	19	42	100	100	100
At. rad. (Golds.)	1·614	1·582	1·569	1·521	1·44	1·372	1·44
M.P. of solute (°C)	630	232	155	321	1,063	1,555	960

Whatever the process of diffusion, the migration of an atom will require first the loosening of that atom from its position of equilibrium in the lattice. It is then natural to attempt to relate diffusion rates and activation energies to the melting points or sublimation energies of the solute and solvent, for these quantities are measures of the work required to loosen atoms from their lattice positions.

Table V. Diffusion of Metals in Gold[76]

Metal	Ni	Pd	Pt	Au
D at 900°C $\times 10^{10}$ ..	6·0	2·3	0·61	—
Q (kcal)	—	37·4	39	51
At. rad. (Golds.)	1·24	1·37	1·39	1·44
M.P. (°C)	1,452	1,555	1,755	1,063

With alloys it is found that an increase in the concentration of one component is attended by an increase in the diffusion coefficient of that component if it leads also to a decrease in the melting point of the alloy, and to a decrease in the diffusion coefficient if it leads to an increase in the melting point, although again other factors operate which prevent a complete correlation.

There is no direct connection between the diffusion rate and the melting point of the solute except that a high diffusion rate is more likely the more the melting points of solute and solvent differ. However,

Table VI. Diffusion of Metals in Copper[76]

Metal	Zn	Al	Si	Sn	Cu[65, 103, 130]
D at 750°C × 10^{10}	1·4	1·6	2·0	2·1	*c* 0·13
Q (kcal)	38	38	40	45	*c* 50
Max. solid sol. (atomic %)	32	16	11	7·5	100
At. rad. (Golds.)	1·37	1·40	1·1	1·58	1·28
M.P. of solute	419	660	1,420	232	1,083

it has sometimes been stated that the activation energy is proportional to the melting point of the solute. Thus Jost[33] shows this to be true for a few elements diffusing in gold. As demonstrated in *Table VIII.*

Table VII. The Direction of Fastest Diffusion in Various Alloys

System	*Min. Distance of Approach in Å*		*Direction of Fastest Diffusion**
Cu-Pt	Cu 2·54	Pt 2·78	Cu *into* Pt
Cu-Zn	Cu 2·54	Zn 2·67 and 2·92	Cu *into* Zn
Fe-Ag	Fe 2·54	Ag 2·88	Fe *into* Ag
Au-Pb	Au 2·88	Pb 3·48	Au *into* Pb
Fe-C	Fe 2·54	C 1·5	C *into* Fe
Pt-Au	Pt 2·78	Au 2·88	Pt *into* Au
Ni-Cu	Ni 2·48	Cu 2·54	Ni *into* Cu
Ni-Au	Ni 2·48	Au 2·88	Ni *into* Au
Cu-Al	Cu 2·54	Al 2·86	Cu *into* Al

* In column 3 the first mentioned metal is of course present in dilute solution

The constancy of Q/T_m in these cases is very interesting, but is by no means universal, as can be seen by examination of the data for lead and silver (*Tables III* and *IV*). Q actually decreases with increase in T_m for the alloys of lead.

There is a more regular correlation between rates of diffusion and the melting point of the solvent, for usually the diffusion rate for any given solute atom is smaller the greater the melting point of the solvent in which it is diffusing, although this is not invariably so. In some cases there is a close proportionality between Q and T_m with the ratio Q/T_m equal to about 21.

This is very nearly equal to the ratio Q/T_m (solute) obtained in *Table VIII* for the activation energies of different solutes diffusing in the same solvent. This equality is rather remarkable. Seith[112] states that Q/T_m (solvent) $\simeq$ 21 is only true for transition metals and metals of the copper group, but the rule is more restricted even than that for it requires that the activation energies for a series of elements diffusing in the same element be equal, and this is not true for copper, palladium, and platinum diffusing in gold, for example.

Table VIII

The Solute Values of Q/T_m for Diffusion in Gold

Solute	*M.P.* (T_m) °*K*	Q/T_m
Cu *diffusing in* Au	1,356	20·2
Pd *diffusing in* Au	1,826	20·5
Pt *diffusing in* Au	2,047	19·0

On the basis of Lindemann's theory of melting,[122] Braune[123] and Van Liempt[124] have derived a relation between the activation energy and the melting point of the diffusing system

Table IX

The Solvent Values of Q/T_m for Various Systems

System	*M.P. of Solvent* (T_m) °*K*	Q/T_m
Cu *in* Au	1,335	20·5
Cu *in* Ni	1,728	20·6
Mo *in* W	3,743	21·3
Au *in* Ag	1,233	21·5

$$Q = 3b^2RT_m \qquad \ldots\ (70)$$

where R is the gas constant and b^2 a constant which varies from metal to metal but is usually $\simeq$ 2. Equation 70 gives values of Q which, while of the correct order of magnitude, are usually much lower than the observed values.

At any one temperature the rate of self diffusion in a metal is generally smaller the higher its melting point. Johnson[107] has pointed out that the activation energy for self diffusion is fairly closely proportional to the melting point temperature T_m and to the sublimation energy E, and *Table X* shows the extent to which these proportionalities hold.

Q/T_m is fairly close to the mean value of 38 and Q/E to the mean value of 0·64 for all the cubic metals considered (except γ iron) and for one of the two Q values (corresponding to the two crystallographic

directions) for the non-cubic metals Zn and Bi. It is, however, impossible to tell, as Johnson and also Steigman, Shockley and Nix[65] concluded, to which of the two quantities, T_m or E, Q is the more closely related. It would help, of course, if there were better agreement between the values of Q measured for the same metal by different observers.

Table X. Johnson's[107] and Additional Figures showing the Proportionality of Activation Energy to Melting Point and Sublimation Energy

Metal	Q (*kcal*)	T_m, °K	E	Q/T_m	Q/E	*Reference*
Cu	57·2	1,356	81·2	42	0·70	65
	61·4			45	0·76	103
Polycryst.	45·1			33	0·56	111
Single cryst.	49·0			36	0·60	
Au	51·0	1,336	92·0	38	0·55	106
	62·9			47	0·68	104
	53·0			40	0·58	132
Pb	28·05	600	47·5	47	0·59	—
Ag	45·9	1,234	68·0	37	0·68	107
Fe α	77·2	1,800	—	41·8	—	131
γ	48·0			26·0	—	—
Bi *perp. to c axis* ..	31·0	544	47·8	57	0·65	8
par. to c axis ..	140			257	2·92	
Zn *perp. to c axis* ..	31·0	693	27·4	45	1·13	9
par. to c axis ..	20·4			29	0·74	
Mean value (cubic metals)	—	—		38	0·64	—

Finally, a most interesting observation, well illustrated in the above tables of data for diffusion in lead, gold, silver, copper, is that the activation energy for diffusion of any metal A in dilute solution in another metal B is always less than the activation energy for self diffusion in B, and usually the more so the more dissimilar the two metals A and B.

Johnson[125] has pointed out that this observation is at variance with the predictions of the simple hole mechanism that we have described above. The diffusion coefficient for such a process we found to be:

$$D = C\,e^{-(E_A + U_0)/RT} = C\,e^{-Q_A/RT} \quad \ldots\ (71)$$

where C is a constant and Q the activation energy is the sum of E_A, the activation energy required for a solute atom A to move into a hole

and U_0 the energy of formation of the hole in the lattice. For self diffusion $E = E_0$ and

$$D_{self} = C\,e^{-(E_0+U_0)/RT} = C\,e^{-Q_0/RT} \quad \ldots\ (72)$$

Johnson explains that this process leading to equations 71 and 72 is really only valid for self diffusing systems or for systems in which $E_A > E_0$, for only then are the successive migrations of any particular solute atom independent of one another. But this requires that $Q_A > Q_0$ which is a result not found experimentally. The case $E_A > E_0$ is then of no practical importance. When $E_A < E_0$ an A atom will be more likely than a B atom to move into an adjacent hole, so that an A atom will interchange position with an adjacent hole a number of times before the hole moves away again by exchanging with the B atom. The successive moves of A atoms will not now be independent as assumed in deriving equation 71, so that the latter is no longer valid. Furthermore the rate of diffusion of A atoms is limited by the rate at which holes diffuse to them through the solvent lattice of B atoms. This is a rate which is proportional to the number of holes and to their diffusion coefficient and is therefore controlled by the activation energy Q_0 so that the activation energy for the diffusion of A atoms can never be less than Q_0 even when $E_A = 0$.

We must then either postulate another mechanism as responsible for the observed diffusions, or if we wish to retain the concept of holes as responsible for diffusion, modify considerably the simple mechanism we have described.

In pursuance of this second alternative Johnson suggests that in certain circumstances a hole and a solute atom might form a relatively stable hole-atom 'molecule' and that the predominant mechanism of diffusion is now the random migration of these 'molecules' through the lattice by the rotation of the hole about the solute atom in the manner of *Figure 17* with the solute atom no longer having to wait for a hole to diffuse to it from the body of the solvent lattice. He derives an approximate relation for the coefficient of diffusion by such a mechanism, which is of the form:

$$D = C\,e^{-(W+S)/RT} = C\,e^{-Q^*/RT}$$

where W is the energy required to form a hole adjacent to a solute atom and S is the activation energy to move a solvent atom into an adjacent hole when this is adjacent also to a solute atom. Since W and

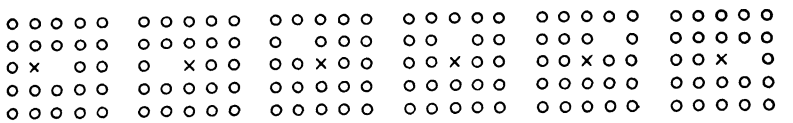


Figure 17. Johnson's diffusion model

S are no longer directly related to E_0, E_A, or U_0 there is no limitation placed on the value of Q^* and it is quite reasonable to have a value of Q^* less than Q_0. This process has also been discussed in more detail by WYLLIE.[126]

REFERENCES

[1] HALLOCK, W. *Z. phys. Chem.* 2 (1888) 378
[2] FARADAY, M. and STODART, J. *Experimental Researches in Chemistry* London, 1859, p 57; *Quart. J. Sci.* 9 (1820) 319
[3] ROBERTS-AUSTEN, W. C. *Phil. Trans. roy. Soc. A* 187 (1896) 383; *Proc. roy. Soc. A* 59 (1896) 281
[4] MEHL, R. F. *Trans. Amer. Inst. min. met. Engrs* 122 (1936) 11
[5] —— *J. appl. Phys.* 12 (1941) 302
[6] —— *Trans. Amer. Inst. min. met. Engrs* 156 (1944) 325
[7] FICK, A. *Ann. Phys. Lpz.* 94 (1855) 59
[8] SEITH, W. *Z. Elektrochem.* 39 (1933) 538
[9] MILLER, P. H. and BANKS, F. R. *Phys. Rev.* 61 (1942) 648
[10] BRICK, R. M. and PHILLIPS, A. *Trans. Amer. Inst. min. met. Engrs* 124 (1937) 331
[11] BOLTZMANN, L. *Ann. Phys. Lpz.* 53 (1894) 959
[12] MATANO, C. *Jap. J. Phys.* 8 (1933) 109
[13] GRUBE, G. and JEDELE, A. *Z. Elektrochem.* 38 (1932) 799
[14] JEDELE, A. *Ibid* 39 (1933) 691
[15] MEHL, R. F., RHINES, F. N. and VAN DER STEINEN, K. A. *Metals and Alloys* 13 (1941) 41
[16] ULLER, K. *Phys. Z.* 32 (1931) 892
[17] DARKEN, L. S. *Trans. Amer. Inst. min. met. Engrs* 150 (1942) 157
[18] BIRCHENALL, C. E. and MEHL, R. F. *Ibid* 171 (1947) 143
[19] GLASSTONE, S., LAIDLER, K. J. and EYRING, H. *Theory of Rate Processes* New York, 1941
[20] ONSAGER, L. and FUOSS, R. M. *J. phys. Chem.* 36 (1932) 2689
[21] DEHLINGER, U. *Z. Phys.* 102 (1936) 633
[22] LAMM, O. *Arkiv. Kemi. mineral. Geol.* 17A, No. 9 (1943)
[23] PFEIL, L. B. *J. Iron Steel Inst.* 119 (1929) 501; DARKEN, L. S. Discussion of reference 26, *Metals Tech.* 14, No. 4 (1947)
[24] BARRER, R. M. *Diffusion in and through solids* Cambridge, 1941
[25] WAGNER, C. *Z. phys. Chem. B* 21 (1933) 25
[26] SMIGELSKAS, A. D. and KIRKENDALL, E. O. *Trans. Amer. Inst. min. met. Engrs* 171 (1947) 130
[27] HERMAN, M. R. Discussion of reference 26, *Ibid* 171 (1947) 140
[28] STEARN, A. E., IRISH, E. M. and EYRING, H. *J. phys. Chem.* 44 (1940) 981
[29] BECKER, R. *Z. Metallk.* 29 (1937) 245
[30] DANIEL, V. *Proc. roy. Soc. A* 192 (1948) 575
[31] KONOBEEVSKY, S. T. *J. Exp. Th. Phy. (U.S.S.R.)* 13 (1932) 185; *ibid* 13 (1943) 418; *J. sci. Instrum.* 24 (1947) 15

[33] EINSTEIN, A. *Ann. Phys. Lpz.* 17 (1905) 549
[34] JOST, W. *Diff. und Chem. reaktionen in festen stoffen* Steinkopf, 1937, p 10
[35] FISHER, J. C., HOLLOMAN, J. H. and TURNBULL, D. *Metals Tech.* 15, No. 2 (1948)
[36] DARKEN, L. S. Discussion of reference 18, *Metals Tech.* 14, No. 5 (1947)
[37] JOHNSON, W. A. *Trans. Amer. Inst. min. met. Engrs* 147 (1942) 331
[38] DARKEN, L. S. *Metals. Tech.* 15, No. 1 (1948)
[39] LE CLAIRE, A. D. Discussion of reference 37, *Metals Tech.* 15, No. 5 (1948)
[40] WAGNER, C. and ENGELHARDT, F. *Z. phys. Chem. A* 159 (1932) 241
[41] SMOLUCHOWSKI, R. *Phys. Rev.* 62 (1942) 539
[42] —— *Ibid* 63 (1943) 438
[43] WELLS, C. and MEHL, R. F. *Trans. Amer. Inst. min. met. Engrs* 145 (1941) 315, 329
[44] BUNGARDT, W. and BOLLENRATH, F. *Z. Metallk.* 30 (1938) 377
[45] GERZRICKEN, S. D., GELLER, A. and TROFIMENKO, V. *Izv. Fiz-Khim Analiza* 16 (1946) 174
[46] LOEB, L. B. *Fundamental processes of electrical discharge in gases* New York, 1939
[47] HOUDREMONT, E. and SCHRADER, H. *Arch. Eisenhütten.* 8 (1935) 445
[48] HENSEL, F. R. *Metals Tech.* 12 (1945) 7
[49] GERZRICKEN, S. D. and DEGHTYAR, I. *Zhur. Tech. Fiz. (U.S.S.R.)* 17 (1947) 881
[50] ELEY, D. D. *Trans. Faraday Soc.* 39 (1943) 168
[51] STEARN, A. E. and EYRING, H. *J. phys. Chem.* 44 (1940) 955
[52] BARRER, R. M. *Trans. Faraday Soc.* 38 (1942) 78
[53] CHANDRASEKHAR, S. *Rev. Mod. Phys.* 15 (1943) 1
[54] BARRER, R. M. *Trans. Faraday Soc.* 37 (1941) 590
[55] BRADLEY, R. S. *Ibid* 33 (1937) 1185
[56] CICHOKI, J. *J. Phys. et Rad.* 7 (1936) 420
[57] WAGNER, C. and SCHOTTKY, W. *Z. phys. Chem.* 11 (1930) 163
[58] MOTT, N. F. and GURNEY, R. W. *Electronic processes in ionic crystals* Oxford, 1940
[59] FRENKEL, J. *Z. Phys.* 35 (1926) 652
[60] WAGNER, C. *Z. phys. Chem.* 38 *B* (1938) 325
[61] SMEKAL, A. *Handbuch der Physik* 24 ii (1933) 880
[62] SMITH, C. S. Discussion of reference 26, *Metals Tech.* 14, No. 4 (1947); *Trans. Amer. Inst. min. met. Engrs* 171 (1947) 136
[63] ROSENHAIN, W. *J. Inst. Met.* 30 (1923) 3
[64] BERNAL, J. D. *Rep. Internat. Conf. on Physics* 2 (1934) 119
[65] HUNTINGTON, H. B. and SEITZ, F. *Phys. Rev.* 61 (1942) 315
[66] STEIGMAN, J., SHOCKLEY, W. and NIX, F. C. *Ibid* 56 (1939) 13
[67] HUNTINGTON, H. B. *Ibid* 61 (1942) 325
[68] HOPKINS, M. *Proc. phys. Soc.* 50 (1938) 703
[69] AWBERY, J. H. *Ibid* 48 (1936) 118
[70] BARRER, R. M. *Ibid* 58 (1946) 321
[71] HARRIS, F. E. *Metals Tech.* 14, No. 5 (1947)
[72] TING-SUI KÊ. *J. appl. Phys.* 19 (1948) 285; *Phys. Rev.* 73 (1948) 267
[73] SMOLUCHOWSKI, R. Discussion of reference 18, *Metals Tech.* 14, No. 5 (1947); *Trans. Amer. Inst. min. met. Engrs* 171 (1947) 159
[74] RHINES, F. N. and MEHL, R. F. *Trans. Amer. Inst. min. met. Engrs* 128 (1938) 185
[75] HARGREAVES, R. *J. Inst. Met.* 64 (1939) 115
[76] JOHNSON, W. A. *Trans. Amer. Inst. min. met. Engrs* 166 (1946) 114
[77] MEHL, R. F. *J. appl. Phys.* 8 (1937) 174
[78] LANGMUIR, I. *J. Franklin Inst.* 217 (1934) 543
[79] SEITH, W. and PERETTI, E. *Z. Elektrochem.* 42 (1936) 570
[80] FRECHE, H. R. *Trans. Amer. Inst. min. met. Engrs* 122 (1936) 324
[81] BEERWALD, A. *Z. Elektrochem.* 45 (1939) 789

[81] KIRKENDALL, E. O., THOMASSEN, L. and UPTHEGROVE, C. *Trans. Amer. Inst. min. met. Engrs* 133 (1939) 186
[82] KUBASCHEWSKI, O. and EBERT, H. *Z. Elektrochem.* 50 (1944) 138
[83] PAIĆ, M. *C. R. Acad. Sci., Paris* 220 (1945) 559
[84] SULLY, A. H. *J. sci. Instrum.* 22 (1945) 245
[85] BÜCKLE, H. *Z. Metallk.* 34 (1942) 130
[86] —— *Z. Elektrochem.* 49 (1943) 238
[87] DUNN, J. S. *J. chem. Soc.* 129 (1926) 2973
[88] GERZRICKEN, S. D., FAINGOLD, M. A., ILKEVICH, G. and SAKHAROV, I. *J. Tech. Phys. U.S.S.R.* 10 (1940) 786
[89] —— and GOLUBENKO, Z. *Izvest. Sekt. Fiziko-Chim. Anal.* (*U.S.S.R.*) 16 (1946) 167
[90] VAN LIEMPT, J. A. M. *Rec. Trav. chim. Pays-Bas* 64 (1945) 234
[91] FONDA, G. R., YOUNG, A. H. and WALKER, A. *Physics* 4 (1933) 1
[92] BRATTAIN, W. H. and BECKER, J. A. *Phys. Rev.* 43 (1933) 428
[93] BOSWORTH, R. C. L. *Proc. roy. Soc. A* 150 (1935) 58
[94] COLEMAN, H. S. and YEAGLEY, H. L. *Phys. Rev.* 62 (1942) 295; *Trans. Amer. Soc. Met.* 31 No. 1 (1943)
[95] —— —— *Phys. Rev.* 65 (1944) 56
[96] DUMOND, J. and YOUTZ, J. P. *J. appl. Phys.* 11 (1940) 357
[97] BURR, A. A., COLEMAN, H. S. and DAVEY, W. P. *Trans. Amer. Soc. Met.* 33 (1944) 73
[98] TANAKA, S. and MATANO, C. *Kyoto Coll. Sci. Mem.* 13 (1930) 343
[99] MATANO, C. *Ibid* 16 (1933) 249
[100] CHEVENARD, P. and WACHÉ, X. *C. R. Acad. Sci., Paris* 217 (1943) 691
[101] —— —— *Rev. Mét.* 41 (1944) 353, 389
[102] —— and PORTEVIN, A. *C. R. Acad. Sci., Paris* 207 (1938) 71
[103] ROLLIN, B. V. *Phys. Rev.* 55 (1939) 231
[104] ZAGRUBSKI, A. *Phys. Zeit. U.S.S.R.* 12 (1937) 118
[105] SEITH, W. and KIEL, A. *Z. Metallk.* 25 (1933) 104
[106] MCKAY, H. A. C. *Trans. Faraday Soc.* 34 (1938) 845
[107] JOHNSON, W. A. *Trans. Amer. Inst. min. met. Engrs* 143 (1941) 107
[108] KUCZYNSKI, G. C. *J. appl. Phys.* 19 (3) (1948) 308
[109] WAGNER, C. *Handbuch d. Metallphysik.* Vol. I, pt II
[110] SMITH, R. P. *J. Amer. chem. Soc.* 68 (1946) 1163
[111] MAIER, M. S. and NELSON, H. R. *Trans. Amer. Inst. min. met. Engrs* 147 (1942) 39
[112] SEITH, W. *Diffusion in Metallen* Berlin, 1939
[113] WELLS, C. and MEHL, R. F. *Trans. Amer. Inst. min. met. Engrs* 140 (1940) 279
[114] DUSHMAN, S. and LANGMUIR, I. *Phys. Rev.* 20 (1920) 113
[115] GERZRICKEN, S. D. and DEGHTYAR, I. *J. Tech. Phys. U.S.S.R.* 17 (1947) 871
[116] MATANO, C. *Proc. Phys. Math. Soc. Japan* 15 (1933) 405
[117] BUGAKOV, W. S. and SIROTKIN, B. *J. Tech. Phys. U.S.S.R.* 4 (1937) 537
[118] SEITH, W. and HERMANN, J. *Z. Elektrochem.* 46 (1940) 213
[119] PETRENKO, B. G. and RUBINSTEIN, B. E. *J. phys. Chem. U.S.S.R.* 13 (1939) 508
[120] JOST, W. *Z. phys. Chem.* 21 (1933) 158
[121] SEN, B. N. *C. R. Acad. Sci., Paris* 199 (1934) 1189
[122] LINDEMANN, F. *Phys. Zeits.* 11 (1910) 609
[123] BRAUNE, H. *Z. phys. Chem.* 110 (1924) 147
[124] VAN LIEMPT, J. A. M. *Rec. Trav. chim. Pays-Bas* 51 (1932) 114
[125] JOHNSON, R. P. *Phys. Rev.* 56 (1939) 814
[126] WYLLIE, G. *Proc. phys. Soc.* 59 (1947) 334
[127] BIRCHENALL, C. E. and MEHL, R. F. Reply to discussion on reference 18, *Metals Tech.* 14, No. 5 (1947)
[128] SEITZ, F. *Phys. Rev.* 74 (1948) 1513

[129] BÜCKLE, H. *Metallforsch* 1 (1946) 175
[130] RAYNOR, C. L., THOMASSEN L. and ROUSE, L. J. *Trans. Amer. Soc. Met.* 30 (1942) 313
[131] BIRCHENALL, C. E. and MEHL, R. F. *J. appl. Phys.* 19 (1948) 217
[132] ZAGRUBSKIJ, A. M. *Bull. acad. Sci. U.S.S.R., Phys. series*

AUTHOR INDEX

Ageev, N. V. 75
Alexander, W. O. 233
Allen, N. P. 236, 258, 264, 266, 277, 279, 280
Allso, P. H. 279
Anderson, W. A. 114, 126
Andrade, E. N. da C. 107, 120, 122, 125, 126
Angier, R. P. 294, 305
Archbutt, L. 132, 161
Archer, R. S. 168–170, 173, 176, 179, 233, 279
Asimow, M. 279, 280
Aston, R. L. 138, 161
Awbery, J. H. 349, 377
Axon, H. J. 71, 76
Ayletsin, V. 126

Baeyertz, M. 278, 280
Bain, E. C. 260, 279, 280
Bakarian, P. W. 282, 305
Baldwin, W. M. 294, 305
Bandel, G. 279
Banks, F. R. 376
Bannister, L. C. 126
Bardgett, W. E. 253, 279
Barnes, A. 144, 161
Barrer, R. M. 333, 335, 336, 338, 340, 349, 361, 371, 376, 377
Barrett, C. S. 31, 74, 138, 161, 231, 234, 293, 305
Bassett, W. H. 294, 300, 305
Bauer, O. 305
Beaman, W. W. 76
Bearden, J. H. 76
Beerwald, A. 377
Becker, R. 209, 318, 320, 376
Becker, J. A. 378
Beilby, G. T. 130, 135, 136, 161
Bengough, G. D. 130, 132, 161
Bennek, H. 279
Betty, B. B. 142, 161
Bilby, B. A. 125
Birchenall, C. E. 322, 324, 325, 353, 368, 376–379
Blough. 163
Boas, W. 139, 140, 153, 161, 287, 291, 305
Boegehold, A. L. 247, 279
Bohlen, E. C. 298, 305
Bohner. 182
Bollenrath, F. 329, 377
Boltzmann, L. 307, 308, 347, 351, 376
Bosworth, R. C. L. 357, 378
Bradley, A. J. 18, 49, 74, 75
Bradley, J. C. 294, 300, 301, 305
Bradley, R. S. 335, 377
Bragg, W. L. 80, 89, 92, 95, 108–113, 125, 126
Brattain, W. H. 135, 161, 378
Braune, H. 373, 378
Brick, R. M. 158, 159, 162, 293, 294, 301, 302, 305, 376
Brillouin, M. 130, 161
Brooks, J. G. 269, 279
Brophy, G. R. 269, 270, 272, 279
Brown, W. F. 104, 125
Bückle, H. 315, 364, 366, 378, 379
Bucknall, E. H. 181–183, 202, 233, 279, 280
Bugakov, W. S. 365, 378
Bungardt, W. 329, 377
Burgess, J. M. 79, 80, 83, 92, 93, 105, 107, 125, 136, 161
Burgess, W. G. 107, 112–114, 125, 126
Burghoff, H. L. 298, 305
Burns, J. L. 279
Burr, A. A. 358, 378
Bush, V. 278

Cahn, R. W. 94, 95, 111, 125, 142, 147, 161
Calvet, J. 204, 205, 207, 218, 219, 233
Carpenter, H. C. H. 138, 153, 161
Castelliz, L. 75
Chalmers, B. 97, 125, 138, 140, 142, 147, 150, 154, 161, 162
Chandrasekhar, S. 334, 377
Channel-Evans, K. M. 74
Chaudron, G. 161
Chevenard, P. 182, 358, 378
Christian, J. W. 12, 74
Cichoki, J. 335, 377
Coffinberry, A. S. 74
Cohen, M. 196, 198, 201–204, 213, 214, 233, 234
Coleman, H. S. 357, 358, 378
Comstock, G. F. 270, 280
Cook, M. 293, 294, 297, 300, 301, 303–305
Cooper, W. E. 258, 266, 279
Cottrell, A. H. 77, 125, 126, 161
Crafts, W. 249, 255, 262, 269, 270, 279

Daniel, V. 319, 376
Darken, L. S. 309, 312, 323, 326, 345, 350, 358, 376
Davey, W. P. 159, 162, 358, 378
Dean, G. R. 159, 162
Dearden, J. 279
Deghtyar, I. 330, 357, 364, 377, 378
Dehlinger, U. 75, 77, 125, 313, 317, 321, 352, 368, 376
Delbar, G. 279
Desch, C. H. 157, 158, 162
Digges, T. G. 280
Doering, W. 75
Dollins, C. W. 142, 161
Donoho, C. K. 279
Douglas, A. M. B. 76
Dumond, J. 358, 378
Dunn, J. S. 357, 378
Dushman, S. 364, 378

Ebert, H. 378
Eckel, E. J. 254, 279
Edwards, C. A. 125
Einstein, A. 199, 320, 377
Ekman, W. 15, 74
Elam, C. F. 125, 138, 153, 161, 287
Eley, D. D. 332, 342, 377
Engelhardt, F. 328, 361, 377
Ensminger, G. R. 161
Ewen, D. 131, 132, 152, 153, 161
Ewing, J. H. 125, 130, 161
Eyres, N. R. 279
Eyring, H. 312, 318, 336, 376, 377

Fahrenhorst, W. 305
Faingold, M. A. 357, 378
Faraday, M. 376
Farineau, J. 76
Fenstermacher, W. H. 245, 257, 279
Fetz, E. 126
Fetzer, M. C. 245, 257, 279
Fick, P. 306, 308, 312, 321, 376
Fink, W. L. 184–189, 196, 198, 200, 201, 205, 207, 210–215, 218, 223, 230, 233, 234
Fisher, J. C. 322, 324, 325, 338, 353, 368, 369, 377
Fonda, G. R. 153, 161, 378
Forsythe, P. J. E. 162
Fourier. 306
Fraenkel, W. 170, 233
Frank, F. C. 99–101, 124, 125
Freche, H. R. 377
Freeman, J. R. 233
French, H. J. 244, 245, 250, 278, 280
Frenkel, J. 99, 125, 377
Friedmann, H. 76
Fuoss, R. M. 313, 376

Gayler, M. L. V. 172, 174–177, 183, 189–195, 198, 199, 203, 204, 216, 218–220, 223, 224, 230, 232–234
Garvey, T. M. 270, 271, 280
Gebhardt, E. 75
Geisler, A. H. 231, 234
Geller, A. 329, 377
Gensamer, M. 125
Gerzricken, S. D. 329, 330, 357, 364, 377, 378
Gibbons, D. F. 126
Glasstone, S. 312, 376
Glen, J. 271, 280
Glocker, R. 293, 305
Goldschmidt, V. M. 3, 74
Göler, F. V. 174, 233, 293, 305
Golubenko, Z. 357, 378
Good, W. A. 74
Goria, C. 75
Gough, H. J. 133, 161
Graf, L. 158, 162
Grange, R. A. 260, 270, 271, 280
Greenwood, J. N. 141, 161
Gregory, C. H. 74, 75
Greninger, A. B. 280
Griffiths, W. T. 236, 264, 278, 279
Grossman, M. H. 245, 268–271, 278, 280
Grube, G. 308, 329, 348, 364, 376
Guertier, W. 76
Guinier, A. 118, 126, 204, 205, 207, 218, 219, 224, 225, 230, 233
Gurney, R. W. 338, 340, 377
Gwyer, A. G. C. 175, 233

Hafner, R. H. 269, 279
Hägg, G. 74
Halla, F. 75
Hallock, 376
Hansen. 165, 173
Hanson, D. 126, 140–142, 150, 161, 233
Hardy, H. K. 305

Hargreaves, F. 134, 136, 161
Hargreaves, M. E. 139, 140, 161
Hargreaves, R. 353, 361, 368, 369, 377
Harris, F. E. 349, 377
Hartree, D. R. 279
Hawkes, M. F. 272, 280
Heidenreich, R. D. 102, 125
Hemingway, E. H. 161
Hengstenberg, J. 179, 205, 233
Hensel, F. R. 330, 377
Herman, M. R. 376
Hermann, J. 365, 378
Hevesey. 369
Hills, R. J. 134, 136, 161
Hinde, R. M. 126
Hino, J. 254, 279
Hodge, J. M. 269, 280
Holloman, J. H. 264, 270, 275–277, 279, 280, 322, 324, 338, 353, 368, 377
Honda, K. 171, 177, 233
Honeycombe, R. W. K. 153, 161
Honeyman, A. J. K. 265, 279
Hopkins, M. 349, 377
Houdremont, E. 280, 330, 377
Howald, T. S. 305
Hultgren, R. 74
Hume-Rothery, W. 2, 5, 10, 12, 13, 32, 43, 44, 54, 62, 66, 71, 74–76
Humfrey, J. C. W. 132, 152, 161
Huntington, H. B. 89, 125, 343, 377

Ilkevich, G. 357, 378
Ingham, J. 278
Irish, E. M. 318, 376

Jackson, R. 278
Jacquet, P. 204, 205, 207, 218, 219, 233
Jaffe, L. D. 260, 264, 270, 275–277, 279, 280
Janitsky, E. J. 278, 280
Jaswon, M. A. 125
Jedele, A. 308, 315, 329, 348, 364, 369, 376
Jeffries, Z. 132, 133, 161, 168–170, 173, 176, 179, 233, 280
Jenkins, C. H. M. 181–183, 202, 233, 278
Jenkins, G. C. H. 278
Jetter, L. K. 208–215, 231
Johnson, R. P. 374–376, 378
Johnson, W. A. 326, 328, 354, 361, 364, 373, 377, 378
Jominy, W. E. 247, 250, 278
Jones, F. W. 243, 249, 254, 261, 278, 279
Jones, H. 29, 42, 45, 67, 73, 74, 76
Jost, W. 369, 372, 377, 378

Kaiser, K. 294, 300, 305
Kauzmann, W. 121, 126
Kê, T. S. 144–149, 161
Kiefer, J. M. 260, 279
Kiel, A. 378
King, R. 142, 147, 154, 161, 162
Kirkendall, E. O. 314, 322, 346 376, 377
Klopsch, O. Z. 244, 245, 250, 278
Kochendörfer, A. 107, 125
Koehler, J. S. 83, 85, 89, 103, 106, 119, 125
Kokuba, S. 177, 232
Konno, S. 170, 232
Konobeevsky, S. T. 320, 376
Kontorova, T. 99, 125
Kornilov, I. I. 12, 74
Koster, W. 305
Kramer, I. R. 269, 279
Kubaschewski, O. 75, 378
Kuczynski, G. C. 378
Kurdjumov, G. 278

Lacombe, P. 151, 160–162
Laidler, K. J. 312, 376
Lamm, O. 312, 376
Lamont, J. L. 249, 255, 262, 269, 270, 279
Langmuir, I. 341, 364, 377, 378
Laque, F. L. 281
Lauvent, P. 125, 167
Laves, F. 36, 38, 74, 75
Le Claire, A. D. 326, 377
Leemann, W. G. 57, 75
Lennard-Jones, J. E. 92, 125
Levenson, L. H. 138, 161
Lewin, G. F. 62, 66, 74, 76
Liedholm, C. A. 279
Lindemann, F. 373, 378
Lipson, H. 18
Little, A. T. 75
Liu, Y. H. 125

Loeb, L. B. 377
Lohberg, K. 38, 75
Loria, E. A. 274, 280
Lorig, C. H. 273, 280
Love, A. E. H. 125
Low, J. R. 125
Lyman, T. 273, 280
Lyssak, L. 278

Mabbot, G. W. 74
Maier, M. S. 378
Makarov, E. S. 40, 75
Manning, G. K. 273, 280
Manning, M. F. 76
Mark, H. 125
Martin, D. L. 294, 305
Masing, G. 193, 233
Matano, C. 308, 350, 351, 358, 366, 376, 378
Mathewson, C. H. 287, 297, 305
Matthaes, K. 305
Matyás, R. 76
May, W. 114, 126
McCarthy, D. E. 264, 279
McCulloch, W. W. 279
McKay, H. A. C. 378
McKeown, J. 141, 161
McLean, D. 249, 279
McMullan, O. W. 280
McQuaid, H. W. 280
Mehl, R. F. 114, 126, 199, 208–215, 231, 234, 272, 274, 279, 280, 306, 312, 319, 322–325, 353, 364, 366, 369, 376–379
Meissner, K. L. 233
Merica, P. D. 163–165, 167–169, 178–181, 195, 200, 208, 213, 214, 233, 234
Metcalfe, G. J. 162
Miller, A. J. 269, 270, 272, 279
Miller, J. V. 126
Miller, P. H. 376
Moeller, K. 74, 75
Moore, H. F. 142, 161
Moore, T. L. 278
Morand, M. 76
Mott, N. F. 42, 45, 73, 81, 102, 115, 116, 119–123, 125, 126, 148, 155, 156, 161, 194, 199, 211, 231, 233, 338, 340, 377
Mueller, F. H. 254, 279
Müller, H. G. 294, 305

Nabarro, F. R. N. 81, 84, 89–91, 98, 102, 115, 116, 119–123, 125, 126
Nadai, A. 125
Nelson, H. R. 378
Neumayr, S. 74
Nielson, H. 144, 161
Nix, F. C. 343, 374
Northcott, L. 249, 279, 283, 305
Norton, M. R. 188, 264, 279
Nye, J. F. 80, 90, 125

Oliver, D. A. 278
O'Neil, H. 279
Onsager, L. 313, 376
Orehoski, M. A. 269, 280
Orowan, E. 77, 87, 99, 111, 118, 121–123, 125, 126
Owen, E. A. 5, 62–64, 74, 76

Paić, M. 378
Palmer, E. W. 305
Parkhouse, R. 216, 234
Patton, W. G. 278, 280
Pauling, L. 3, 46, 53, 74, 75
Peierls, R. 90, 125
Peretti, E. 369, 377
Petch, N. J. 278
Petrenko, B. G. 365, 378
Pfeil, L. B. 49, 75, 236, 264, 278, 279, 376
Phillips, A. 158, 159, 162, 214, 234, 376
Phillips, D. 125
Phragmén, G. 74
Pickus, M. R. 287, 297, 305
Polanyi, M. 77, 125
Polyak, E. V. 146, 161
Portevin, A. 182, 358, 378
Post, C. B. 245, 257, 278
Potaszkin, R. 279
Powell. 274
Prandtl, L. 77, 125
Pratt, J. N. 75
Preston, G. D. 174, 175, 177, 183, 189, 204, 207, 220, 222, 224, 233, 234
Pumphrey, W. I. 243, 249, 254, 261, 279

Quincke, G. 130, 161
Quinney, H. 104

Rawdon, W. 161
Raynor, G. V. 15, 43, 49, 74–76, 379
Read, T. A. 106, 111, 121, 125, 126

Reynolds, P. W. 15, 62, 66, 74, 76
Rhines, F. N. 188, 325, 353, 364, 366, 369, 376, 377
Richards, T. Ll. 293, 294, 297, 303, 305
Roberts, E. A. O'D. 74
Roberts, E. W. 74, 76
Roberts-Austen, W. C. 306, 376
Röhner, F. 195, 226–230, 234
Rollin, B. V. 378
Rosbaud, P. 125
Roscoe, R. 107, 125
Rosenhain, R. 125, 126, 130–135, 141, 152, 153, 161, 170, 173, 176, 177, 184, 233, 341, 377
Rosenthal, W. 76
Rouse, L. J. 379
Rowlands, V. W. 74
Rubenstein, B. E. 365, 378
Russell, J. E. 260, 279
Russell, T. F. 243, 248, 249, 278

Sachs, G. 174, 175, 233, 293, 297, 305
Sakarov, I. 357, 378
Sandee, J. 114, 116
Sarjant, R. J. 278
Scheil, E. 260, 279
Scheuer, E. 170, 233
Schmid, E. 125, 198, 205, 284, 287, 291, 305
Schottky, W. 377
Schrader, H. 280, 330, 377
Schultze, G. E. R. 75
Scott, H. 161, 163–165, 167, 168, 171, 200, 233
Seith, W. 362, 365, 369, 373, 376–378
Seitz, F. 106, 121, 125, 328, 338, 343, 344, 377, 378
Sen, B. N. 370, 378
Sergeev, S. V. 146, 161
Shapiro, C. L. 278
Shephard, H. D. 280
Shepherd, B. F. 256, 279
Shockley, W. 102, 105, 343, 374, 377
Shternfeld, A. 75
Shuttleworth, R. 154, 161
Siegel, S. 269, 280
Sirotkin, B. 365, 378
Smekal, A. 340, 377
Smigelskas, A. D. 314, 322, 324, 346, 358, 359, 368, 376
Smith, C. L. 122, 123, 126
Smith, C. S. 305, 340, 377
Smith, D. W. 184–189, 196, 198, 200, 201, 205, 207, 210–215, 218, 223, 230, 233, 234
Smith, R. P. 323, 361, 368, 378
Smoluchowski, R. 271, 280, 328, 377
Snoek, J. L. 125
Spretnak, J. 298, 305
Stearn, A. E. 318, 336, 376, 377
Steigman, J. 343, 374, 377
Steinberg, S. 279
Steven, W. 279
Stockdale, D. 233
Stodart, J. 376
Strauss, J. 278
Strawbridge, D. J. 75
Sully, A. H. 378
Sutton, H. 233
Sykes, W. P. 280

Tammann, G. 178, 233
Tanabe, T. 233
Tanaka, S. 358, 378
Taylor, A. 18, 49, 74, 75, 305
Taylor, G. I. 77, 78, 82, 86, 92, 102, 104–107, 110, 125, 149, 161, 287, 288, 305
Tengenér, S. 75
Thomas, D. E. 283, 305
Thomassen, L. 378, 379
Timpe, A. 82, 125
Ting-Sui Kè 377
Townsend, J. R. 134, 161
Trofimenko, V. 329, 377
Troiano, A. R. 273, 280
Turnbull, D. 322, 324, 338, 353, 368, 377
Tyndall, E. P. T. 111, 126

Uller, K. 308, 376
Upthegrove, C. 378
Urban, S. F. 278, 280

Vacher, H. C. 287, 288, 305
Van der Steinen, 364, 376
Van Liempt, J. A. M. 351, 357, 364, 366, 373, 378
Van Winkle, D. 144, 161
Vegard, L. 61, 73, 76
Venturello, G. 75
Volmer, M. 193, 208, 233
Volterra, V. 82, 125

Wachié, X. 358, 378
Wagner, C. 328, 360, 361, 376–378
Wagstaff, J. B. 278
Wakeman, D. W. 75, 76
Waldron, M. B. 75, 76
Walker, H. L. 254, 278
Wallbaum, H. J. 75
Waltenberg, R. G. 163–165, 167, 168, 171, 200, 233
Wassermann, G. 179, 185, 198, 219, 233, 283, 305
Webb, W. 74
Weber, A. 193, 208, 233
Weerts, G. 185, 219, 233
Weerts, J. 291, 302, 305
Weinman, E. W. 278
Wells, C. 323, 364, 368, 377, 378
Westgren, A. 12, 74, 76
Westlinning, H. 75
Wever, F. 239, 278, 287, 305
Wheeler, M. A. 126, 140–142, 150, 161
Williams, E. C. 304
Williamson, J. C. 248, 279
Willstrop, J. W. 233
Wilm, A. 163, 233
Wilson, F. H. 293, 294, 301, 302, 305
Witte, H. 15, 21, 36, 38, 74, 75
Womer, H. K. 141, 161
Wood, W. A. 110, 161
Wooster, W. A. 283, 305
Wrazej, W. J. 238
Wyllie, G. 376, 378

Yannequis, N. 151, 160–162
Yeagley, H. L. 357, 378
Youtz, J. P. 358, 378
Young, A. H. 378

Zachariasen, W. H. 3, 74
Zagrubski, A. 378
Zagrubskij, A. M. 379
Zaharova, M. 75
Zener, C. 12, 29–31, 74, 111, 126, 143–146, 161, 240, 279
Zintl, E. 74

SUBJECT INDEX

Absorption of electrons by transition elements 45, 53
Activities of components in diffusing systems 312, 313, 321, 322, 325, 330, 351,
Additivity law (Vegard) 61, 73 [360 *et seq.*
Age hardening 163 *et seq.*
influence of size factors 11
Ageing 163 *et seq.*
aluminium-magnesium alloy 184, 215
-silver alloy 231
-zinc alloy 172, 173
colloidal theory 173
copper-iron alloy 184, 212
-phosphorus alloy 184
dislocations, theory of 110 *et seq.*, 195
early theories of 165 *et seq.*
effect of cooling rate from solution treatment 221 *et seq.*
See also Quenching
plastic deformation 110 *et seq.*, 181, 188, 196 *et seq.*, 225
quenching 163, 164, 196 *et seq.*
on density 180, 186, 187, 199, 224
ductility 187
electrical resistance 170, 171, 176, 187, 194, 199
hardness. *See* Age hardening
tensile strength 164, 182, 184
yield strength 187
expansion on 182, 201 *et seq.*
Gayler theory 223–226
Guinier-Preston zones 204 *et seq.*, 211 *et seq.*, 230, 231
intergranular corrosion on 201
knot theory 178 *et seq.*, 187, 189, 195, 196, 201
lattice alteration theory 226 *et seq.*
changes 168, 170, 173, 174, 176 *et seq.*, 181, 184, 186, 187, 190, 195,
distortion theories 176 *et seq.* [201, 203, 213, 224, 230, 231
microstructural investigation of 166, 167, 204 *et seq.*, 216 *et seq.*
over-ageing 114, 118
precipitation mechanism 168, 208 *et seq.*
theory 166, 167, 172 *et seq.*
rate 181, 188, 196 *et seq.*
retrogression effects 177, 193, 200, 226
segregate zone theory 182
slip interference theory 169, 179
solid solubility, importance 165, 166
sub-zero 166
temperature considerations 163, 164, 183, 190 *et seq.*, 197, 198, 201, 202, 204
theories of, colloidal 173
dislocation 110 *et seq.*
early 165 *et seq.*
Gayler 223 *et seq.*
knot 178 *et seq.*, 187, 189, 195, 196, 201
Jeffries-Archer 169, 179
lattice alteration 226 *et seq.*
distortion 176 *et seq.*
precipitation 166, 167, 172 *et seq.*, 187
Rohner 226 *et seq.*
segregate zone 182
three stage 201 *et seq.*
two stage 189 *et seq.*

Ageing theta phase (Al-Cu alloys) 185 *et seq.*
three stage theory 201 *et seq.*
two stage theory 189, 190
worked metals 110 *et seq.*
x-ray investigation 174, 175, 179, 184, 185, 205, 206, 219, 224, 225
Aluminium, anelastic effects 144
creep 141
crystal boundaries, melting 151
face-centred cubic structure 171
in steel, effect on austenite range 239
hardenability 272
intercrystalline attack 160
metallic diffusion in 372
recrystallization 114
texture 293
slip bands in 107
stress, interaction of grains under 138
viscosity 146
Aluminium alloys, metallic diffusion in 366, 367
bronze, diffusion in 314, 341
Aluminium-copper alloys, ageing of. *See* Ageing
crystal boundaries, compositions 158, 159
theta phase 185 *et seq.*
Aluminium-iron-nickel alloys, constitution 49
Aluminium-magnesium alloys, ageing 184, 215
boundary precipitation in 160
metallic diffusion in 329
Aluminium-silver alloy, ageing 231
Aluminium-zinc alloys, equilibrium diagram 173
ageing 172, 173
Amorphous cement. *See* Crystal boundaries
Anelastic effects in metals 110, 111, 143, 144
Annealed metals, theory of dislocations 97, 110 *et seq.*
strip and wire, directional properties 300 *et seq.*
preferred orientation 292 *et seq.*
recrystallization texture 292 *et seq.*
Anisotropy and metallic diffusion 307
cupping test for strip 300 *et seq.*, 304
in annealed material 300 *et seq.*
plastically deformed material 289 *et seq.*
Anti-isomorphism in fluospar structures 42, 44
Antimony-tin alloys, thermal etching 153
Atomic size, effect in ternary alloys 55, 56
on alloy formation 2, 3
solid solution formation 5, 7
measures of 2, 3
significance in face-centred cubic structures 32
Atoms, surface migration 154
Austempering 237, 267
Austenite, effects of alloying elements 239, 240
lattice 238
solubility of carbon in 235, 239
transformations on cooling 241
Axial ratio and electron concentration in magnesium alloys 69, 70

Bainite 275
Bainitic hardenability 275, 276

Bismuth, metallic diffusion in 307, 374
Blue-brittleness of iron 97
Body-centred cubic phase structure in electron compounds 14, 15
Boron, effect on transformations in steel 274, 275
Boundary precipitation 158, 160
slip 143, 147 *et seq.*
Brass, beta, intercrystalline attack 158
lattice structure 15
eta, lattice structure 21, 22
gamma, lattice structure 21
metallic diffusion in 309, 314, 321 322, 325, 329, 341, 357, 369
strip, anisotropy in 291
preferred orientation in. *See* Preferred orientation
recrystallization texture 294, 298, 299
rolling texture 290
Brillouin zones effect of temperature 28 *et seq.*
in body-centred cubic structure 25
cobalt-aluminium compounds 53
copper, gold and silver 26
face-centred cubic structure 25
fluorspar structure 43
magnesium 27, 28, 67 *et seq.*
nickel-aluminium compound 53
nickel arsenide structure 41

CADMIUM, solid solutions in 12
thermal hardening of 111
Cast irons, hardening of 267
Chromium in steel, effect on austenite range 239
bainite formation 275, 276
eutectoid composition 240
hardenability 268, 269, 271, 273, 276
Chromium-molybdenum steels, fatigue strength of 264
hardenability of 276
isothermal transformations in 273, 274
Close-packed hexagonal structure, lattice distortion in 67 *et seq.*
copper, gold and silver 17
copper-zinc system 70
magnesium 27, 67
nickel-arsenide type 39, 41
Closest approach, distance of, as measure of atomic size 3
Cobalt in steel, effect on austenite range 239
diffusion of carbon 329
hardenability 268, 271, 272
Cold-worked state, density of dislocations 104, 105, 110
lattice rotation 105
structure 135
Cold-working, ageing after. *See* Plastic deformation
energy absorbed 104, 110
Colloidal theory of ageing 173
Columnar structure 128, 129
Compound formation in magnesium alloys 10
Continuous cooling, transformations of steel on 259 *et seq.*
Copper in iron, effect on carbon diffusion 330
steel, effect on austenite range 239
bainite formation 275

Copper metallic diffusion in 369, 372, 374
preferred orientation in. *See* Preferred orientation
recrystallization textures in. *See* Recrystallization textures
self-diffusion in 343, 344, 359
strip, effect of cold rolling and annealing on 282 *et seq.*
thermal etching 152
Copper alloys, metallic diffusion in 364–366, 368
Copper-aluminium alloys, compound, ageing effect. *See* Age-hardening
structure of 33
equilibrium diagram 57, 165
metallic diffusion in 366, 368, 372
Copper-aluminium-nickel alloys 60
beta-phase 19
Copper-beryllium alloys, age-hardening 210
boundary precipitation in 160
metallic diffusion in 366, 368
Copper-iron alloys 184, 212
Copper-nickel alloys, metallic diffusion in 365, 372
recrystallization textures 294
Copper-nickel-tin alloys 61
Copper-phosphorus alloys, ageing 184
Copper-silicon alloys, metallic diffusion in 366, 368
with cobalt and nickel 60
Copper-silver alloys, ageing 210
Copper-tin alloys, equilibrium diagram 57
metallic diffusion in 366, 368
Copper-tin-magnesium alloys 61
Copper-zinc alloys (*see also* Brass), equilibrium diagram 57
metallic diffusion in 366, 368, 372
Corrosion. *See* Intercrystalline corrosion
Cracking, fatigue 236
quenching 237, 267, 277
welding 267
Creep, fracture in 141
importance of crystal boundaries 141
intercrystalline fissures in 142
micro-creep 97
quasi-viscous 120, 121
relative rotation of crystals in 142
resistance of ferritic steels 267
tertiary 142, 150
torsional, under constant stress 144, 145
transient 120 *et seq.*
viscous component 150
Critical cooling rate for hardening of steels 242
Critical diameter. *See also* Hardenability 242, 244
Critical dispersion size in precipitation hardening 114, 118
Crystallites, boundaries between, and the theory of dislocations 92 *et seq*, 107 [110, 112, 124
centres of recrystallization 113, 114
in cold-worked material 110, 135
lower limiting size 110
rotation 107, 112, 136
slip in 108
theory of crystal boundaries 133
Crystal boundaries 127 *et seq.*
amorphous cement theory 130 *et seq.*, 141, 151
angles between 154–156
chemical composition 158 *et seq.*
chemical effects at 157 *et seq.*

Crystal boundaries columnar type 128, 129
cracks (fissures) at 141, 142, 150
crystallite theory 133
diffusion at 157
energy of 127, 136, 155, 157
equiaxial type 128, 129
fatigue, effect of 134
fracture at 141, 150
grooves at. *See* Thermal etching
importance in creep 141
mechanical properties and effects of 130 *et seq.*
melting 150 *et seq.*
movement (migration) 154 *et seq.*
mutual interaction of crystals across 138 *et seq.*
origin 127 *et seq.*
precipitation at 158
slip at 143, 147 *et seq.*
stress relaxation across 147
theories of, amorphous cement 130 *et seq.*, 141, 151
crystallite 133
transitional lattice 134, 137, 141, 151, 155–157, 160
thermal fluctuations, effect of 149, 160
viscous behaviour 144–146, 148, 152, 153
widening 134
Crystals, boundaries of. *See* Crystal boundaries
columnar, formation 128, 129
equiaxial, formation 128, 129
fragmentation 92, 105
growth of, in recrystallization 113
nuclei for 124
stimulated 114
lattices of. *See* Lattices
mosaic blocks in 92, 106, 136
plastic strain 106, 107
preferred orientation of. *See* Preferred orientation
single, yield strength of 138
slip in. *See* Slip
strain energy 108, 156
twin boundaries of 151, 160
Cupping test for detection of anisotropy in strip 309 *et seq.*, 304

Deep drawing, importance of preferred orientation 281
Density changes on ageing 180, 186, 187, 199, 224
Differential analyser in thermal calculations 245
Diffusion of metals in metals 306 *et seq.*
activities of components in 312, 313, 321, 322, 325,
anisotropy in 307, 344 [330, 351, 360 *et seq.*
chemical potentials of components 310, 311
coefficients, determination 345 *et seq.*
experimental results 362 *et seq.*
interpretation 321 *et seq.*
effect of impurities 330
lattice defects 336 *et seq.*, 340, 341
mosaic blocks 340
grain boundary 312, 319, 328, 355, 363
in relation to ageing 189, 199

Diffusion of metals in metals Kirkendall effect 316
mechanisms, atomic interchange 335 *et seq.*
'hole' or 'vacancy' 336 *et seq.*
interstitial 312, 339 *et seq.*
involving one atom 312 *et seq.*
involving two atoms 315 *et seq.*
operation of several 318 *et seq.*
mobilities of components 314, 316, 321, 322, 325, 330 331 *et seq.*
phenomenological theory 310 *et seq.*
self 326, 327, 343, 344
determination of coefficients 354, 359 *et seq.*
ternary and higher systems 325 *et seq.*
uphill 319 *et seq.*, 330
Directional properties, cupping test 300
in cold-rolled copper and brass strip 290, 291
drawn copper wire 291
hot worked or extruded aluminium 292
Dislocation lines 80, 82, 132, 136
pairs, activation energy for formation 88, 89
formed in crystallites 109
on slip planes 87, 91
strain energy of 85, 87
strength 80, 83
Dislocations 77 *et seq.*
assemblies 92 *et seq.*, 106, 107, 110
behaviour in thermal softening 112, 113
coalescence 87
compound 80
cores of 83, 84, 89
definition 78
density 82
in annealed metals 97
cold worked state 104, 105, 110
in relation to solute atoms 96, 97
slip planes 101
dissolution 110, 112, 113
edge 77, 79, 80, 82, 85, 93, 95
effect of thermal fluctuations 107, 108, 119, 123
on diffusion 330
energy 90, 104
fast 98 *et seq.*
flexible 114 *et seq.*
relation to ageing and annealing 110 *et seq.*, 195
interaction 85 *et seq.*
energy 85, 86, 94
forces 111
forming pairs. *See* Dislocation pairs
origin 123, 124
positive and negative 79, 85, 103
properties 80 *et seq.*
mobility (movement) 80 *et seq.*, 90, 102, 112
resultant adjacent half crystal displacement 81
screw 79, 80, 83, 93, 100, 101, 136
segregation of solute atoms round 95 *et seq.*
self energies 85, 86
shear stress required to move 90 *et seq.*
strain energy 84, 85
stresses round 80, 82 *et seq.*

Dislocations types of 77 *et seq.*
compound 80
edge 77, 79, 80, 82, 85, 93, 95, 96, 100
positive and negative 79, 85, 103
screw 79, 80, 83, 93, 100, 101, 136
width 90, 91
Drawing, preferred orientation after 285, 286
Ductility of annealed strip, directional values for 302
effect of ageing on 187
Duralumin, ageing of. *See* Ageing and Age hardening
preferred orientation after drawing 285
Dynamic rigidity 144–147

Elastic anisotropy in body-centred cubic structures 30
Electrical resistance, effect of ageing 170, 171, 176, 180, 187, 193 *et seq.*, 199, 201 *et seq.*, 224, 226
Electrochemical character and electron compound formation 17
solid solution formation 3, 4
effect on solubility in magnesium 10
factor in solid solutions (high valency solvent) 12
Electron absorption in alloys of the transition metals 49, 53
Electron/atom ratio, effect on solubility of zinc and silver in aluminium 12
in aluminium and transition metal compounds 49, 51
copper-aluminium-nickel alloys 19
nickel-aluminium alloys 18
saturated solutions in copper and silver 5, 10
magnesium 11, 12
relationship to homogeneity 16
phase structure 15 *et seq.*
fluorspar type 43
Laves phase type 36, 37
nickel-arsenide type 40
zinc blende type 44
wurtzite type 44 *et seq.*
significance in solid solution formation 4 *et seq.*, 9 *et seq.*
Electron band theory 45
in alloys of transition and high valency metals 53
Electron compounds, definition 13
in system silver-zinc-antimony 58
types, electron/atom ratio 3 : 2 13
7 : 4 21, 15
21 : 13 15, 19
beta brass (B.C.C.) 15
eta brass (F.C.C.) 21, 22
gamma brass 21
Electron concentration and axial ratio for magnesium alloys 69, 70
Electron distribution in transition metals 46 *et seq.*
Electron kinetics 23
Electronic theory of alloys 1, 23 *et seq.*
Electronic structure of transition metals 45
Electrostatic forces and shear resistance in lattices 67
Elongation, effect of ageing on. *See* Ductility
Equicohesive temperature 133
Eutectoid temperature, effect of size factor in ternary systems 57
Exhaustion hardening 102, 103

FACE-centred cubic structure, dependence on atomic size 32
development of rolling texture in 287 *et seq.*
of aluminium 71
recrystallization texture in 294
Fatigue, changes on slip planes during 133
cracking in steel forgings 236
Ferrite 235, 240, 241, 247
Fissures, intercrystalline 142, 143, 150
Fluorspar structure 42
Fractional time theory for transformations in steels 260
Fracture, intercrystalline 132, 141
transcrystalline 131, 132, 141
Free energy curves for ternary alloys 59, 60
determination of equilibrium 59
principles of alloy formation 3, 4
Full zone compounds 44

GAMMA loop 239
Gayler theory of ageing 223 *et seq.*
wave mechanical interpretation 194
Glide. *See* Slip
Gold, metallic diffusion in 369, 371–374
recrystallization texture in 293
self-diffusion in 359
Gold-copper alloys, metallic diffusion in 336, 358, 373
Gold-nickel alloys, metallic diffusion in 308, 315, 364, 365, 371, 372
Gold-palladium alloys, metallic diffusion in 308, 315, 336, 364, 365, 371, 373
Gold-platinum alloys, metallic diffusion in 308, 315, 336, 364, 365, 371–373
Grain boundaries. *See* Crystal boundaries
Grain growth. *See* Crystals
Grains. *See* Crystals
Guinier-Preston zones of ageing theory 204 *et seq.*, 211 *et seq.*, 230, 231

HARDENABILITY, bainitic 275, 276
bands 248
critical diameter, a measure of 242, 244
effect of alloying elements 268 *et seq.*
grain size 268, 275
segregation 253
ideal critical diameter, a measure of 246, 249, 268–270
Jominy end quench test for measuring 247 *et seq.*
measurement of, critical diameter, a basis for 242, 244
early methods 245, 246
ideal critical diameter, a basis for 246, 249, 268–270
in deep hardening steels 257 *et seq.*
in shallow hardening steels 256 *et seq.*
Jominy end quench test for 247 *et seq.*
Shepherd test for 256, 257
transformation data, a basis for 257 *et seq.*
pearlitic 276
Shepherd test for measuring 256, 257
systematizing of, data for 267 *et seq.*
transformation data, a basis for measuring 257 *et seq.*
variations of in bars and billets 251 *et seq.*

Hardening, age. *See* Age hardening
of steels 235 *et seq.*
critical cooling rate 242
of hyper-eutectoid composition 235
mechanical properties after 262 *et seq.*
precipitation. *See* Precipitation hardening
strain. *See* Strain hardening
thermal 111
Heteropolarity of nickel-arsenide structure 39, 41
Homogeneity, effect on phase structure of electron compounds 17
range for aluminium-magnesium-copper system 51
-nickel system 51
-zinc system 51
nickel-arsenide type structure 39
silver alloys, relationship to size factor in 58

IDEAL critical diameter 246, 249, 268–270
Indium, effect on diffusion of silver into thallium 331
Interatomic distance, measure of atomic size 2
Intercrystalline attack 158, 160
of beta-brass by mercury 158
corrosion 157
on ageing 201
cracks (fissures) 142, 150
fracture 132, 141
penetration 158
stress corrosion of silver-gold alloys 158
Intermediate transformations in steels 236, 237, 248, 258, 259, 263–265, 267, 273–275]
Intermetallic compounds, $CuAl_2$ structure type 33
Cu_2In, structure of 41
fluorspar structure type 43
Laves phase types 34 *et seq.*
Ni_2In, structure of 41
ternary alloys, formation in 54
Interpenetrating lattice, formation of on ageing 179
Internal friction 111, 131, 144, 145, 147
Internal stresses and the theory of dislocations 82, 102
Interstitial metal/non-metal compounds 32
Ionic crystals, diffusion in 315
interaction in alloy formation 9
size and solid solution formation 3, 9
Iron, blue-brittleness of 97
deformation at elevated temperatures 132
diffusion of carbon in 309–311, 322, 323, 349, 364, 368, 374
, effect of other metals on 328–330
effect on ageing of aluminium-copper alloys 181
lattice structure 238
solubility of added elements in 239
strain ageing 98, 110, 111
thermal etching 152
yield point 98
Iron-nickel alloys, recrystallization textures in 300
Iron-nickel-copper alloys, recrystallization textures in 294
Iron-tungsten alloys, age-hardening 210
Isomorphy of intermetallic compounds 38, 39, 49
Isothermal transformations in steels 237, 257–261, 267, 275

JEFFRIES-ARCHER slip interference theory of ageing 169, 179
Jominy end quench test for measuring hardenability 247 *et seq.*

KEYING action on slip planes in ageing 169
Kirkendall effect 315
Knot theory of ageing 178 *et seq.*, 187, 189, 195, 196, 201

LATTICE alteration theory of ageing (Rohner) 226 *et seq.*
changes on ageing 168 *et seq.*, 173, 174, 176 *et seq.*, 181, 184, 186, 190,
dislocations. *See* Dislocations [195, 201, 203, 213, 224, 230, 231
distortion, effect of solute valency 62, 63, 65, 66, 69, 71
volume 64, 65
factor in solid solution formation 5
in aluminium alloys 71 *et seq.*
copper, silver and gold alloys 62 *et seq.*
magnesium alloys 67 *et seq.*, 73
ternary alloys 55
Vegard's additivity law 61, 62, 73
elastic strains in 136
parameter, effect of ageing on 184, 185, 187, 201, 202, 224
in ternary alloys 55
rotation in cold working 105
spacings of alloys 61 *et seq.*
, calculation of 61, 62
aluminium alloys 71 *et seq.*
copper, silver and gold alloys 61 *et seq.*
magnesium alloys 67 *et seq.*
transitional. *See* Crystal boundaries, theories of
Laves phases (AB_2) 34 *et seq.*
factors controlling structure of 36
Lead creep, rotation of crystals 142
crystal boundaries of 129
metallic diffusion in 306, 365, 370, 372, 374
self-diffusion in 359
Lead-antimony alloys, crystal boundaries 134

MAGNESIUM, alloys, lattice distortion 67 *et seq.*
Brillouin zones of 27, 67 *et seq.*
close-packed hexagonal structure 27
effect on ageing rate of aluminium-copper alloys 181
solid solutions in 10
Magnesium-copper compound $MgCu_2$. *See* Laves phases
Magnesium-nickel compound $MgNi_2$. *See* Laves phases
Magnesium-silicide, effect on ageing 167, 168, 171, 172, 175, 182
Magnesium-zinc compound $MgZn_2$. *See* Laves phases
Manganese in iron, effect on carbon diffusion 329
steel, effect on austenite range 338
bainite formation 275, 276
eutectoid composition 240
hardenability 268–271, 274, 276

Manganese-molybdenum steels, properties of, 263–266
Martempering 267
Martensite, formation 236, 237
structure 238, 239
Mechanical properties, directionality in annealed strip 301, 303, 304
Mercury-lead alloys, diffusion in 330, 357
Microstructural investigation of ageing 166, 167, 204 *et seq.*, 216 *et seq.*, 225
Mobilities of components in diffusing systems 314, 316, 321, 322, 325, 330–335
Molybdenum in iron, effect on carbon diffusion 329
steel, effect on austenite range 239
bainite formation 275
eutectoid composition 240
hardenability 271
Mosaic blocks (structure) and metallic diffusion 330, 340
theory of dislocations 92, 106, 136

NICKEL in iron, effect on carbon diffusion 329
steel, effect on austenite range 239
bainite formation 275, 276
eutectoid composition 240
hardenability 268, 270, 271, 276
recrystallization texture in 294
Nickel-aluminium alloys, electron/atom ratio importance 18, 19
Nickel-arsenide structure, (type AB) 39
relationship to other types 41
Nickel-chromium steels, fatigue strength 241, 264, 273, 276
hardenability 276
maximum transformation velocity 272, 273
Nickel-chromium-molybdenum steels, fatigue strength 264
hardenability 276
isothermal transformations 259, 260
Nickel-chromium-molybdenum-vanadium steels, intermediate transformations 274
Nickel-copper alloys, metallic diffusion in 308, 329, 358, 364, 365, 368, 373
Nickel-gold alloys, metallic diffusion in 321
Nickel-indium compound, structure of 41
Nickel-iron alloys, recrystallization texture in 294
Nickel-molybdenum steels, fatigue strength 264
Nitrogen in steel, effect on austenite range 239
Nucleation theory (Volmer and Weber) 208, 209
Nuclei, crystal growth around 124, 127
recrystallization centres, effect as 113 *et seq.*
solidification around 127

ORIENTATION—*see* Preferred orientation
Overageing 114, 118

PAULING's theory of the transition metals 46
Pearlite 236, 244, 258, 259, 273
Pearlitic hardenability 275, 276
Phosphorus in steel, effect on austenite range 239

Planck's constant 23
Plastic deformation and the theory of dislocations 77 *et seq.*
effect on ageing 181, 188, 196 *et seq.*
preferred orientation after 284
Platinum, metallic diffusion in 373
Polishing, surface flow on 130
Polygonization 94, 111
Precipitation ageing theory 166, 167, 172 *et seq.*, 185, 187
at crystal boundaries 158, 160
hardening and the dislocation theory 114 *et seq.*
mechanism in ageing 168, 208 *et seq.*
Preferred orientation in non-ferrous metals 281 *et seq.*
annealed material 292 *et seq.*
copper-base alloy, cold rolled strip 285 *et seq.*
castings 283
drawn wire 284, 285
electrodeposits 283
anisotropy after 289, 292
causes 281
occurrence 283
plastic deformation after 284
progressive development of 288, 289
x-ray investigation of 281 *et seq.*

Quenching cracking 237, 267, 277
effect on ageing 163, 164, 196 *et seq.*
intensity 243
stages 244

Recovery after annealing 110
Recrystallization 111, 113, 114, 129, 133
development 294 *et seq.*
in annealed wire and strip 293
textures 292 *et seq.*
Relative valency effect 12
Relaxation. *See* Stress relaxation
Resoftening 114, 118
Retrogression effects on ageing 177, 193, 220, 226
Rigidity. *See* Dynamic rigidity
Rohner's theory of ageing 226 *et seq.*
Rolling texture, theory 288 *et seq.*

Scaling 314
Schottky lattice defects 337 *et seq.*
S-curves, and hardenability 258, 259
determination 257
effect of carbide formers on 275
for nickel-chromium-molybdenum steel (En110) 259
Season cracking 132
Segregate zone theory of ageing 182

Segregation, effect on hardenability 253
Shepherd hardenability test 256, 257
Silicon in steel, effects on austenite range 239
bainite formation 275
hardenability 268
Silicon-iron, recrystallization texture 300
Silver, metallic diffusion in 365, 369, 371–374
self-diffusion 359
thermal etching 152, 154
Silver alloys, metallic diffusion in 315, 326, 327, 365, 371–373
Silver-cadmium-antimony system 61
Silver-copper alloys, ageing 196
Silver-gold alloys, intercrystalline stress corrosion 158
metallic diffusion in 371, 373
Silver-thallium antifriction alloys, metallic diffusion in 330
Silver-zinc alloys, metallic diffusion in 315
Silver-zinc-antimony alloys, ageing 57, 58, 61
Single crystals, yield strength 138
Size factor, effect on eutectoid temperature in ternary systems 57
phase structure of electron compounds 15
solid solubility 2, 5, 6, 9
in high valency metals 12
in magnesium 10
Slip (glide), activation energy for nucleating 109
bands, formation of and the theory of dislocations 98 *et seq.*, 107, 124
near crystal boundaries 132
changes in strain energy in crystals, due to 108, 109
crystal boundary 143, 147, 148, 150
dislocations, in relation to 77 *et seq.*
interference theory of ageing (Jeffries and Archer) 169, 179
internal friction opposing 131
planes, changes on in fatigue 133
keying action on ageing 169, 178
precipitation on 97
resistance on ageing 178
Sodium-thallium structure 32
Solid solubility, low value for gold, explanation 8, 9
of copper in aluminium and age-hardening 165, 166
relationship to valency 4 *et seq.*, 10 *et seq.*
Solid solutions, formation of in relation to electrochemical character 3, 4
electron/atom ration 4
ionic size 3, 9
valency 4 *et seq.*, 10 *et seq.*
in aluminium 8, 12
cadmium 12
copper, gold, and silver 2 *et seq.*
magnesium 10
zinc 7, 12
Solute valency, effect on phase structure of electron compounds 17
solid solubility 6, 7
Steel, austempering 237, 267
austenite 235, 238–241
bainite 275, 276
bainitic hardenability 275, 276
cracking in, fatigue 236
quench 237, 267, 277
weld 267
effects of aluminium 239, 272
chromium 239, 240, 264, 268, 269, 271–276

Steel, effects of cobalt 239, 268, 271, 272, 330
copper 239, 275, 331
manganese 239, 240, 268–272, 274–276
molybdenum 239, 240, 271, 272, 275
nickel 239–241, 259, 260, 264, 268, 270–276
nitrogen 239
phosphorus 239
silicon 239, 268, 275
sulphur 271
titanium 239, 249, 270
tungsten 239, 240
vanadium 271, 272
ferrite 235, 240, 241, 247
hardenability. *See* Hardenability
hardening. *See* Hardening
hyper-eutectoid 235
martempering 267
martensite 236–239
mechanical properties after incomplete hardening 262 *et seq.*
pearlite 244, 258, 259, 273
pearlitic hardenability 275, 276
quenching 237, 242 *et seq.*, 247 *et seq.*, 267, 277
strain ageing 98
tempering 237, 262 *et seq.*
transformations. *See* Transformations
welding, cracking 267
yield point 98
Strain ageing 98, 110
Strain hardening during ageing 195
exhaustion 102, 103
structural 102 *et seq.*, 123
Stress relaxation 144, 146, 147
Stress-strain curves for cubic structure metals 106
Strip manufacture, importance of anisotropy 301
Structural strain hardening and the theory of dislocations 102 *et seq.*, 123
Substitutional intermetallic compounds 32
Sub-zero ageing 166
Sulphur in steel, effect on hardenability 271
Supercooling, effect on crystallization 128
Surface effects in relation to crystal boundaries 152 *et seq.*
Surface migration of atoms 154
Surface tension forces and ageing 168
Swaging, preferred orientation after 284

TARNISHING 315
Tempering of steel, effect on mechanical properties 237, 262 *et seq.*
Tensile strength, effect of ageing 164, 182, 184
Ternary alloys, effect of compound formation 54
electron/atom ratio, influence of 56
eutectoid temperature and size factor relationship 57
lattice distortion 55
metallic diffusion in 326 *et seq.*
equilibrium 54
Tetrahedral structures in Laves phases 35
zinc blende and wurtzite 44
Thermal etching of metals 152 *et seq.*

Thermal hardening in cadmium and zinc 111
Thermal softening, and the theory of dislocations 111 *et seq.*
Theta phase in aluminium-copper alloys 185 *et seq.*
Three stage theory of ageing 201 *et seq.*
Tin, creep, micro- 97
rotation of crystals in 142
crystal boundaries, angles between 154, 155
formation of 129
melting of 150, 151
shear of 147
Titanium in iron, effect on diffusion of carbon 331
steel, effect on austenite range 239
hardenability 249, 270
Torsional creep 144, 145
Torsion, anelastic effects in 144
Transcrystalline fractures 131, 132, 141
Transformations in steel, continuous cooling, on 259 *et seq.*
fractional time theory of 260
intermediate 236, 237, 248, 258, 259, 263–267, 273–275
isothermal 237, 257–261, 267, 275
velocity of 258, 259, 272–274
Transitional lattice theory of crystal boundaries 134, 137, 141, 151, 155–157, 160
Transition metals, alloys with high valency metals 53
behaviour in aluminium alloys 45
electron distribution 48
Nickel-arsenide structure with electronegative elements 39, 45
Tungsten in iron, effect on diffusion of carbon 329
steel, effect on austenite range 239
eutectoid composition 240
metallic diffusion in 357, 363, 373
thermal etching 153
Twin crystals, boundaries of 151, 160
Two stage theory of ageing 189, 190

VALENCY, definition 1
effects in copper, gold and silver alloys 62
solid solubility, relationship to 4 *et seq.*, 10 *et seq.*
Vanadium in steel, effect on hardenability 271, 272
Vegard's additivity law 61, 73

WAVE mechanical interpretation of Gayler ageing theory 194, 195
Welding, cold 306
cracking in 267
Widmanstatten mechanisms 199
Wire drawing, preferred orientation after 284, 285
Wurtzite structure 44

YIELD point of iron 98
Yield strength, effect of ageing 187, 195
crystal boundaries 138
in relation to dislocation theory 102, 105, 111, 118
precipitation hardening 114, 118
Young's modulus, directionality in annealed strip 302

Zinc, anelastic effects 111
creep 121
crystal boundaries 129
crystallite transition surfaces 95
metallic diffusion 307, 372, 374
self-diffusion 344, 359
solid solutions 7, 12
thermal etching 152
thermal hardening 159
Zinc blende structure 44
Zinc-copper alloys, compositions of crystal boundaries in 159